Numer. Math. (2012) 120:189–229
DOI 10.1007/s00211-011-0411-2

Numerische Mathematik

Analysis and comparison of numerical methods for the Klein–Gordon equation in the nonrelativistic limit regime

Weizhu Bao · **Xuanchun Dong**

Received: 18 August 2010 / Revised: 15 September 2010 / Published online: 11 September 2011

Abstract We analyze rigourously error estimates and compare numerically temporal/spatial resolution of various numerical methods for solving the Klein–Gordon (KG) equation in the nonrelativistic limit regime, involving a small parameter $0 < \varepsilon \ll 1$ which is inversely proportional to the speed of light. In this regime, the solution is highly oscillating in time, i.e. there are propagating waves with wavelength of $O(\varepsilon^2)$ and $O(1)$ in time and space, respectively. We begin with four frequently used finite difference time domain (FDTD) methods and obtain their rigorous error estimates in the nonrelativistic limit regime by paying particularly attention to how error bounds depend explicitly on mesh size h and time step τ as well as the small parameter ε. Based on the error bounds, in order to compute 'correct' solutions when $0 < \varepsilon \ll 1$, the four FDTD methods share the same ε-scalability: $\tau = O(\varepsilon^3)$. Then we propose new numerical methods by using either Fourier pseudospectral or finite difference approximation for spatial derivatives combined with the Gautschi-type exponential integrator for temporal derivatives to discretize the KG equation. The new methods are unconditionally stable and their ε-scalability is improved to $\tau = O(1)$ and $\tau = O(\varepsilon^2)$ for linear and nonlinear KG equations, respectively, when $0 < \varepsilon \ll 1$. Numerical results are reported to support our error estimates.

Mathematics Subject Classification (2000) 35L70 · 65N06 · 65N12 · 65N15 · 65N35 · 81-08

W. Bao (✉) · X. Dong
Department of Mathematics and Center for Computational Science and Engineering, National University of Singapore, Singapore 119076, Singapore
e-mail: bao@math.nus.edu.sg
Url: http://www.math.nus.edu.sg/∼bao/

X. Dong
e-mail: dong.xuanchun@nus.edu.sg

1 Introduction

Consider the dimensionless relativistic Klein–Gordon (KG) equation in d-dimensions ($d = 1, 2, 3$) [31–33]

$$\varepsilon^2 \partial_{tt} u - \Delta u + \frac{1}{\varepsilon^2} u + f(u) = 0, \quad \mathbf{x} \in \mathbb{R}^d, \quad t > 0, \tag{1.1a}$$

with initial conditions given as

$$u(\mathbf{x}, 0) = \phi(\mathbf{x}), \quad \partial_t u(\mathbf{x}, 0) = \frac{1}{\varepsilon^2} \gamma(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \tag{1.1b}$$

Here $u = u(\mathbf{x}, t)$ is a real-valued field, $\varepsilon > 0$ is a dimensionless parameter which is inversely proportional to the speed of light [31–33], ϕ and γ are given real-valued functions, $f(u)$ is a dimensionless real-valued function independent of ε and satisfies $f(0) = 0$. In practice, the typical nonlinearity is the pure power case, i.e. $f(u) = \lambda u^{p+1}$ with $p \geq 0$ and $\lambda \in \mathbb{R}$ [9,10,17–20,31–34,40,42,44,47]. In fact, the above KG equation is also known as the relativistic version of the Schrödinger equation under proper non-dimensionalization [31–33] and it is used to describe the motion of a spinless particle (see, e.g. [13,41], for its derivation). The KG equation (1.1) is time symmetric or time reversible. In addition, if $u(\cdot, t) \in H^1(\mathbb{R}^d)$ and $\partial_t u(\cdot, t) \in L^2(\mathbb{R}^d)$, it also conserves the *energy* [31–33], i.e.

$$\begin{aligned} E(t) &:= \int_{\mathbb{R}^d} \left[\varepsilon^2 \left(\partial_t u(\mathbf{x}, t)\right)^2 + |\nabla u(\mathbf{x}, t)|^2 + \frac{1}{\varepsilon^2} u^2(\mathbf{x}, t) + F\left(u(\mathbf{x}, t)\right) \right] \mathrm{d}\mathbf{x} \\ &\equiv \int_{\mathbb{R}^d} \left[\frac{1}{\varepsilon^2} \gamma^2(\mathbf{x}) + |\nabla \phi(\mathbf{x})|^2 + \frac{1}{\varepsilon^2} \phi^2(\mathbf{x}) + F\left(\phi(\mathbf{x})\right) \right] \mathrm{d}\mathbf{x} := E(0), \quad t \geq 0, \end{aligned} \tag{1.2}$$

where

$$F(u) = 2 \int_0^u f(s)\, \mathrm{d}s, \quad u \in \mathbb{R}. \tag{1.3}$$

For fixed $\varepsilon > 0$($O(1)$-speed of light regime), e.g. $\varepsilon = 1$, the KG equation has gained a surge of attention in both analytical and numerical aspects. Along the analytical front, the Cauchy problem was investigated, e.g. in [3,9,17,19,27,29,42,44]. In particular, for the defocusing case (i.e. $F(u) \geq 0$ for $u \in \mathbb{R}$) the global existence of solutions was established in [9], and for the focusing case (i.e. $F(u) \leq 0$ for $u \in \mathbb{R}$) possible finite time blow-up was shown in [3]. For more results in this regime, we refer the readers to [2,10,34,37,40,43,47] and references therein. In the numerical aspect, various numerical schemes were proposed and studied in the literatures.

For instance, standard finite difference time domain (FDTD) methods such as energy conservative, semi-implicit and explicit finite difference discretizations were proposed and analyzed in [1,15,30,39,48]. Other approaches, like finite element or spectral discretization, were also studied in [11,12,14,50]. Comparisons of different methods in this regime were carried out in [28,39] and references therein.

However, in the nonrelativistic limit regime, i.e. if $0 < \varepsilon \ll 1$ or the speed of light goes to infinity, the analysis and efficient computation of the KG equation (1.1) are mathematically rather complicated issues. The analysis difficulty is mainly due to that the energy $E(t)$ in (1.2) becomes unbounded when $\varepsilon \to 0$. Recently, Machihara et al. [32] studied such limit in the energy space, and Masmoudi and Nakanishi [33] analyzed such limit in a strong topology of the energy space. For more recent progresses made on this topic, we refer to [35,36,51]. Their results show that the solution propagates waves with wavelength $O(\varepsilon^2)$ and $O(1)$ in time and space, respectively, when $0 < \varepsilon \ll 1$. On the other hand, this highly oscillatory nature in time provides severe numerical burdens, making the computation in the nonrelativistic limit regime extremely challenging. To our knowledge, so far there are few results on the numerics of the KG equation in this regime.

The aim of this paper is to study the efficiency of the frequently used FDTD methods applied in the nonrelativistic limit regime, to propose new numerical schemes and to compare their resolution capacities in this regime. We start with the detailed analysis on the stability and convergence of four standard implicit/semi-implicit/explicit energy conservative or non-conservative FDTD methods. Here we pay particular attention to how the error bounds depend explicitly on the small parameter ε in addition to the mesh size h and time step τ. Based on the estimates, in order to obtain 'correct' numerical approximations when $0 < \varepsilon \ll 1$, the meshing strategy requirement (ε-scalability) for those frequently used FDTD methods is:

$$\tau = O(\varepsilon^3), \quad h = O(1), \tag{1.4}$$

which suggests that the standard FDTD methods are computationally expensive for the KG equation (1.1) as $0 < \varepsilon \ll 1$. To relax the ε-scalability, we then propose new numerical methods whose ε-scalability is optimal for both time and space in view of the inherent oscillatory nature. The key ideas of the new schemes are: (i) to apply either Fourier pseudospectral or centered finite difference discretization for spatial derivatives; and (ii) to discretize the highly oscillatory second-order ordinary differential equations (ODEs) in phase space by using the Gautschi-type exponential integrator [16,24] which was well demonstrated in the literatures that it has favorable properties compared to standard time integrators for oscillatory second-order differential equations [21,22,25,26]. For the linear KG equation, the Gautschi-type time integration does not introduce any time discretization error. Rigorous error estimates show that the ε-scalability of the new methods is improved to

$$\tau = O(1), \quad h = O(1), \tag{1.5}$$

for the linear KG equation, and respectively, to

$$\tau = O(\varepsilon^2), \quad h = O(1), \tag{1.6}$$

for the nonlinear KG equation. Thus, the Gautschi-type method offers compelling advantages over commonly used FDTD methods in temporal resolution when $0 < \varepsilon \ll 1$.

The paper is organized as follows. In Sect. 2, four second-order FDTD methods are reviewed and their stability and convergence are analyzed in the nonrelativistic limit regime. In Sect. 3, new numerical methods are proposed and analyzed rigorously. In Sect. 4, numerical comparison results are reported. Finally, some concluding remarks are drawn in Sect. 5. Throughout the paper, we adopt the standard Sobolev spaces and use the notation $p \lesssim q$ to represent that there exists a generic constant C which is independent of h, τ and ε such that $|p| \le C\,q$.

2 FDTD methods and their analysis

In this section, we apply the commonly used FDTD methods to the KG equation (1.1) [15,28,30,39,48] and analyze their stability and convergence in the nonrelativistic limit regime. For simplicity of notations, we shall only present the numerical methods and their analysis in one space dimension (1D). Generalization to higher dimensions is straightforward and results remain valid without modifications. Similar to most works in the literatures for the analysis and computation of the KG equation (cf. [1,11,14,15,28,30,39,48,50] and references therein), in practical computation, we truncate the whole space problem onto an interval $\Omega = (a, b)$ with periodic boundary conditions. In 1D, the KG equation (1.1) with periodic boundary conditions collapses to

$$\varepsilon^2 \partial_{tt} u - \partial_{xx} u + \frac{1}{\varepsilon^2} u + f(u) = 0, \quad x \in \Omega = (a, b), \quad t > 0, \tag{2.1a}$$

$$u(a, t) = u(b, t), \quad \partial_x u(a, t) = \partial_x u(b, t), \quad t \ge 0, \tag{2.1b}$$

$$u(x, 0) = \phi(x), \quad \partial_t u(x, 0) = \frac{1}{\varepsilon^2} \gamma(x), \quad x \in \bar{\Omega} = [a, b]; \tag{2.1c}$$

with $\phi(a) = \phi(b)$, $\phi'(a) = \phi'(b)$, $\gamma(a) = \gamma(b)$ and $\gamma'(a) = \gamma'(b)$.

2.1 FDTD methods

Choose mesh size $h := \Delta x = (b - a)/M$ with M being an even positive integer, time step $\tau := \Delta t > 0$ and denote grid points and time steps as

$$x_j := a + jh, \quad j = 0, 1, \ldots, M; \quad t_n := n\tau, \quad n = 0, 1, 2, \ldots.$$

Let u_j^n be the approximation of $u(x_j, t_n)(j = 0, 1, \dots, M,\ n = 0, 1, \dots)$ and introduce the finite difference discretization operators as

$$\delta_t^+ u_j^n = \frac{u_j^{n+1} - u_j^n}{\tau}, \quad \delta_t^- u_j^n = \frac{u_j^n - u_j^{n-1}}{\tau}, \quad \delta_t^2 u_j^n = \frac{u_j^{n+1} - 2u_j^n + u_j^{n-1}}{\tau^2},$$
$$\delta_x^+ u_j^n = \frac{u_{j+1}^n - u_j^n}{h}, \quad \delta_x^- u_j^n = \frac{u_j^n - u_{j-1}^n}{h}, \quad \delta_x^2 u_j^n = \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{h^2}.$$

It is easy to check that $\delta_t^2 = \delta_t^+\delta_t^- = \delta_t^-\delta_t^+$ and $\delta_x^2 = \delta_x^+\delta_x^- = \delta_x^-\delta_x^+$. Here, we consider four frequently used FDTD methods [15,28,30,39,48] to discretize the problem (2.1): for $j = 0, 1, \dots, M-1,\ n = 1, 2, \dots,$

I. *Implicit energy conservative finite difference* (*Impt-EC-FD*) *method*

$$\varepsilon^2\delta_t^2 u_j^n - \frac{1}{2}\delta_x^2\left(u_j^{n+1} + u_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + G\left(u_j^{n+1}, u_j^{n-1}\right) = 0; \tag{2.2}$$

II. *Semi-implicit energy conservative finite difference* (*SImpt-EC-FD*) *method*

$$\varepsilon^2\delta_t^2 u_j^n - \delta_x^2 u_j^n + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + G\left(u_j^{n+1}, u_j^{n-1}\right) = 0; \tag{2.3}$$

III. *Semi-implicit finite difference* (*SImpt-FD*) *method*

$$\varepsilon^2\delta_t^2 u_j^n - \frac{1}{2}\delta_x^2\left(u_j^{n+1} + u_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + f(u_j^n) = 0; \tag{2.4}$$

IV. *Explicit finite difference* (*Expt-FD*) *method*

$$\varepsilon^2\delta_t^2 u_j^n - \delta_x^2 u_j^n + \frac{1}{\varepsilon^2}u_j^n + f(u_j^n) = 0. \tag{2.5}$$

Here,

$$G(v, w) = \int_0^1 f\left(\theta v + (1-\theta)w\right)\, d\theta = \frac{F(v) - F(w)}{2(v-w)}, \quad \forall\, v, w \in \mathbb{R}, \tag{2.6}$$

with $F(u)$ defined in (1.3). The initial and boundary conditions in (2.1) are discretized as

$$u_0^n = u_M^n,\ u_{-1}^n = u_{M-1}^n, \quad n \ge 0, \quad u_j^0 = \phi(x_j), \quad j = 0, 1, \dots, M, \tag{2.7}$$
$$u_j^1 = \phi(x_j) + \frac{\tau}{\varepsilon^2}\gamma(x_j) + \frac{\tau^2}{2\varepsilon^2}\left[\delta_x^2\phi(x_j) - \frac{1}{\varepsilon^2}\phi(x_j) - f(\phi(x_j))\right]. \tag{2.8}$$

Clearly, the above four FDTD methods are time symmetric or time reversible, i.e. they are unchanged if we interchange $n+1 \leftrightarrow n-1$ and $\tau \leftrightarrow -\tau$. Expt-FD is an explicit method, whereas Impt-EC-FD, SImpt-EC-FD and SImpt-FD are implicit methods. At each time step, SImpt-FD needs to solve a linear coupled system, SImpt-EC-FD needs to solve a nonlinear decoupled system, and Impt-EC-FD needs to solve a fully nonlinear coupled system.

Denote $X_M = \{v = (v_0, v_1, \ldots, v_M) \mid v_0 = v_M\} \subset \mathbb{R}^{M+1}$ and we always use $v_{-1} = v_{M-1}$ for the vector $v \in X_M$ if it is involved. Letting $\{v_j^n,\ j = 0, 1, \ldots, M;\ n = 0, 1, \ldots\}$ be any grid function satisfying $v_0^n = v_M^n (n = 0, 1, \ldots)$ and using $v_{-1}^n = v_{M-1}^n$ if they are involved, thus we have $v^n = (v_0^n, v_1^n, \ldots, v_M^n) \in X_M$ and define its standard discrete l^2 norm, semi-H^1 norm, semi-H^2 norm and l^∞ norm as

$$\|v^n\|_{l^2}^2 = h \sum_{j=0}^{M-1} \left|v_j^n\right|^2, \quad \|\delta_x^+ v^n\|_{l^2}^2 = h \sum_{j=0}^{M-1} |\delta_x^+ v_j^n|^2, \tag{2.9}$$

$$\left\|\delta_x^2 v^n\right\|_{l^2}^2 = h \sum_{j=0}^{M-1} \left|\delta_x^2 v_j^n\right|^2, \quad \|v^n\|_{l^\infty} = \max_{0\le j\le M-1} \left|v_j^n\right|, \quad n \ge 0. \tag{2.10}$$

For the first two methods Impt-EC-FD and SImpt-EC-FD, one can easily show that they conserve the energy in the discretized level, i.e.

Lemma 1 *The method Impt-EC-FD (2.2) conserves the discrete energy as*

$$E^n = \varepsilon^2 \left\|\delta_t^+ u^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^+ u^n\right\|_{l^2}^2 + \left\|\delta_x^+ u^{n+1}\right\|_{l^2}^2\right) + \frac{1}{2\varepsilon^2}\left(\left\|u^n\right\|_{l^2}^2 + \left\|u^{n+1}\right\|_{l^2}^2\right)$$
$$+\frac{h}{2}\sum_{j=0}^{M-1}\left[F(u_j^n) + F\left(u_j^{n+1}\right)\right] \equiv E^0, \quad n = 0, 1, 2, \ldots. \tag{2.11}$$

Similarly, the method SImpt-EC-FD (2.3) conserves the discrete energy as

$$\widetilde{E}^n = \varepsilon^2 \left\|\delta_t^+ u^n\right\|_{l^2}^2 + h \sum_{j=0}^{M-1} \delta_x^+ u_j^n \cdot \delta_x^+ u_j^{n+1} + \frac{1}{2\varepsilon^2}\left(\left\|u^n\right\|_{l^2}^2 + \left\|u^{n+1}\right\|_{l^2}^2\right)$$
$$+\frac{h}{2}\sum_{j=0}^{M-1}\left[F(u_j^n) + F\left(u_j^{n+1}\right)\right] \equiv \widetilde{E}^0, \quad n = 0, 1, 2, \ldots. \tag{2.12}$$

Proof The proof proceeds in the analogous lines as in [30,48] for the standard KG equation, i.e. $\varepsilon = 1$ in (2.1), and we omit the details here for brevity. □

2.2 Stability analysis

By using the standard von Neumann analysis [45], we have the following stability results for the FDTD methods:

Theorem 1 *Suppose* $f(u)$ *is linear, i.e.* $f(u) = \alpha u$ *with* α *a constant satisfying* $\alpha > -\varepsilon^{-2}$*, then we have:*

(i) *The method Impt-EC-FD* (2.2) *is unconditionally stable for any* $\tau > 0$, $h > 0$ *and* $\varepsilon > 0$.

(ii) *When* $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) \le 0$*, the method SImpt-EC-FD* (2.3) *is unconditionally stable for any* $\tau > 0$ *and* $h > 0$*; and when* $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) > 0$*, it is conditionally stable under the stability condition*

$$\tau \le \frac{2h\varepsilon^2}{\sqrt{4\varepsilon^2 - h^2(1+\varepsilon^2\alpha)}}. \tag{2.13}$$

(iii) *When* $-\varepsilon^{-2} < \alpha \le \varepsilon^{-2}$*, the method SImpt-FD* (2.4) *is unconditionally stable for any* $\tau > 0$ *and* $h > 0$*; and when* $\alpha > \varepsilon^{-2}$*, it is conditionally stable under the stability condition*

$$\tau \le \frac{2\varepsilon^2}{\sqrt{\varepsilon^2\alpha - 1}}. \tag{2.14}$$

(iv) *The method Expt-FD* (2.5) *is conditionally stable under the stability condition*

$$\tau \le \frac{2h\varepsilon^2}{\sqrt{4\varepsilon^2 + h^2(1+\alpha\varepsilon^2)}}. \tag{2.15}$$

Proof Noticing $f(u) = \alpha u$, plugging

$$u_j^{n-1} = \sum_l \hat{U}_l e^{2ijl\pi/M}, \quad u_j^n = \sum_l \xi_l \hat{U}_l e^{2ijl\pi/M}, \quad u_j^{n+1} = \sum_l \xi_l^2 \hat{U}_l e^{2ijl\pi/M},$$

into (2.2)–(2.5), with ξ_l the amplification factor of the lth mode in phase space, we obtain the characteristic equation with the following structure

$$\xi_l^2 - 2\theta_l\,\xi_l + 1 = 0, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1, \tag{2.16}$$

where $\theta_l \in \mathbb{R}$ is determined by the corresponding method and may vary for different methods. Solving the above equation, we have $\xi_l = \theta_l \pm \sqrt{\theta_l^2 - 1}$. The stability of numerical schemes amounts to

$$|\xi_l| \le 1 \iff |\theta_l| \le 1, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1. \tag{2.17}$$

(i) For the method Impt-EC-FD (2.2), noticing $\alpha > -\varepsilon^{-2}$, we have

$$0 \le \theta_l = \frac{2\varepsilon^4}{2\varepsilon^4 + \tau^2(\varepsilon^2\lambda_l^2 + \varepsilon^2\alpha + 1)} \le 1, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1, \quad (2.18)$$

with

$$\lambda_l = \frac{2}{h} \sin\left(\frac{l\pi}{M}\right), \quad \mu_l = \frac{2l\pi}{b-a}, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1. \quad (2.19)$$

This implies that the method Impt-EC-FD (2.2) is unconditionally stable for any $\tau > 0$, $h > 0$ and $\varepsilon > 0$.

(ii) For the method SImpt-FD (2.4), we have

$$\theta_l = \frac{2\varepsilon^4 - \tau^2\varepsilon^2\lambda_l^2}{2\varepsilon^4 + \tau^2(\varepsilon^2\alpha + 1)}, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1. \quad (2.20)$$

From (2.19), we see that

$$0 \le \lambda_l^2 \le \frac{4}{h^2}, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1. \quad (2.21)$$

Thus, when $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) \le 0$, or $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) > 0$ with the condition (2.13), for $l = -\frac{M}{2}, \dots, \frac{M}{2} - 1$, we obtain

$$(\varepsilon^2\lambda_l^2 - \varepsilon^2\alpha - 1)\tau^2 \le \left(\frac{4\varepsilon^2}{h^2} - \varepsilon^2\alpha - 1\right)\tau^2 \le 4\varepsilon^4 \Longrightarrow |\theta_l| \le 1.$$

(iii) For the method SImpt-EC-FD (2.3), we have

$$\theta_l = \frac{2\varepsilon^4 - \tau^2\varepsilon^2\alpha}{2\varepsilon^4 + \tau^2(\varepsilon^2\lambda_l^2 + 1)}, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1. \quad (2.22)$$

Noticing (2.21), when $-\varepsilon^{-2} < \alpha \le \varepsilon^{-2}$, or $\alpha > \varepsilon^{-2}$ with the condition (2.14), we get

$$\tau^2(\varepsilon^2\alpha - 1 - \varepsilon^2\lambda_l^2) \le \tau^2(\varepsilon^2\alpha - 1) \le 4\varepsilon^4 \Longrightarrow |\theta_l| \le 1, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1.$$

(iv) For the method Expt-FD (2.5), we have

$$\theta_l = \frac{2\varepsilon^4 - \tau^2(\varepsilon^2\lambda_l^2 + \varepsilon^2\alpha + 1)}{2\varepsilon^4}, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1. \quad (2.23)$$

Combining (2.21) and (2.15), for $l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1$, we have

$$\tau^2(\varepsilon^2\lambda_l^2 + 1 + \varepsilon^2\alpha) \leq \tau^2\left(\frac{4\varepsilon^2}{h^2} + 1 + \varepsilon^2\alpha\right) \leq 4\varepsilon^4 \Longrightarrow |\theta_l| \leq 1.$$

The proof is completed. □

2.3 Error estimates

Motivated by the analytical results in [32,33] for the KG equation, here we make the following assumptions on the exact solution u of (2.1)

$$\text{(A)}\quad u \in C^4([0,T]; W^{1,\infty}) \cap C^3([0,T]; W^{2,\infty}) \cap C^2([0,T]; W^{3,\infty}) \cap C([0,T]; W_p^{5,\infty}),\ \left\|\frac{\partial^{r+s}}{\partial t^r \partial x^s} u(x,t)\right\|_{L^\infty(\Omega_T)} \lesssim \frac{1}{\varepsilon^{2r}},\quad 0 \leq r \leq 4\ \&\ 0 \leq r+s \leq 5$$

where $W_p^{m,\infty} = \{v \in W^{m,\infty} \mid v^{(l)}(a) = v^{(l)}(b), 0 \leq l \leq m-1\}$ for $m \geq 1$, $\Omega_T = \Omega \times [0,T]$ and $0 < T < T^*$ with T^* the maximum existence time of the solution; and assumption on the function $f(v)$ in (2.1)

$$\text{(B1)}\quad \|f'(v)\|_{L^\infty(\mathbb{R})} + \|f''(v)\|_{L^\infty(\mathbb{R})} \lesssim 1, \quad \text{or}$$
$$\text{(B2)}\quad \|f'(v)\|_{L^\infty(\mathbb{R})} + \|f''(v)\|_{L^\infty(\mathbb{R})} + \|f'''(v)\|_{L^\infty(\mathbb{R})} \lesssim 1.$$

Define the grid 'error' function $e^n \in X_M (n \geq 0)$ as

$$e_j^n = u(x_j, t_n) - u_j^n, \quad j = 0, 1, \ldots, M, \quad n = 0, 1, 2, \ldots, \tag{2.24}$$

with u_j^n the approximations obtained from FDTD methods.

For the method Impt-EC-FD (2.2), we can establish the following error estimate (see detailed proof in Appendix I below):

Theorem 2 *Assume* $\tau \lesssim h$ *and under assumptions* (*A*) *and* (*B2*), *there exist constants* $\tau_0 > 0$ *and* $h_0 > 0$ *sufficiently small and independent of* ε *such that, for any* $0 < \varepsilon \leq 1$, *when* $0 < \tau \leq \tau_0$ *and* $0 < h \leq h_0$, *we have the following error estimate for the method Impt-EC-FD* (2.2) *with* (2.7) *and* (2.8)

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{2.25}$$

For Expt-FD method (2.5), we assume the stability condition

$$\tau \leq \min\left\{\frac{\varepsilon h}{2}, \frac{\varepsilon^2}{\sqrt{2}}, \frac{2h\varepsilon^2}{\sqrt{4\varepsilon^2 + h^2(1 + \varepsilon^2\|f'\|_{L^\infty(\mathbb{R})})}}\right\}, \tag{2.26}$$

and can establish the following error estimate (see detailed proof in Appendix II below):

Theorem 3 *Assume $\tau \lesssim h$ and under assumptions (A) and (B1), there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < \tau \le \tau_0$ and $0 < h \le h_0$ and under the stability condition (2.26), we have the following error estimate for the method Expt-FD (2.5) with (2.7) and (2.8)*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}. \tag{2.27}$$

Similarly, for the methods SImpt-EC-FD (2.3) and SImpt-FD (2.4), we have

Theorem 4 *Assume $\tau \lesssim h$ and under assumptions (A) and (B2), there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < \tau \le \tau_0$ and $0 < h \le h_0$ and under the stability condition $\tau \le \varepsilon h/\sqrt{2}$, we have the following error estimate for the method SImpt-EC-FD (2.3) with (2.7) and (2.8)*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}. \tag{2.28}$$

Proof Follow the analogous proofs to Theorems 2 and 3 and we omit the details here for brevity. □

Theorem 5 *Assume $\tau \lesssim h$ and under assumptions (A) and (B1), there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < \tau \le \tau_0$ and $0 < h \le h_0$, we have the following error estimate for the method SImpt-FD (2.4) with (2.7) and (2.8)*

$$\left\|e^n\right\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}. \tag{2.29}$$

Proof Follow the analogous proofs to Theorems 2 and 3 and we omit the details here for brevity. □

Based on Theorems 2–5, the four FDTD methods studied here share the same temporal/spatial resolution capacity in the nonrelativistic limit regime. In fact, given an accuracy bound $\delta > 0$, the ε-scalability of the four FDTD methods is:

$$\tau = O\left(\varepsilon^3\sqrt{\delta}\right) = O(\varepsilon^3), \quad h = O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1. \tag{2.30}$$

Remark 1 We can establish the same error estimates in the theorems under much weaker assumption on the nonlinear function f, i.e. relax the assumption (B1) to $f \in C^2(\mathbb{R})$, and respectively, (B2) to $f \in C^3(\mathbb{R})$. This can be done by replacing the assumption $\tau \lesssim h$ in Theorems 2–5 by $\tau \lesssim \min\{h,\ \varepsilon^3\sqrt{C_d(h)}\}$, and combing the method of mathematical induction (see [4,5] and the proof of Theorem 9 below) for

SImpt-FD and Expt-FD, and respectively, a cut-off function technique to deal with the nonlinear function f [4,49] for Impt-EC-FD and SImpt-EC-FD, with the following inverse inequality [4,49]

$$\|u^n\|_{l^\infty} \lesssim \frac{1}{C_d(h)}\left[\|\delta_x^+ u^n\|_{l^2} + \|u^n\|_{l^2}\right], \quad C_d(h) = \begin{cases} 1, & d=1, \\ 1/|\ln h|, & d=2, \\ h^{1/2}, & d=3. \end{cases} \tag{2.31}$$

3 New efficient numerical methods and their analysis

In this section, we propose new numerical methods which have better temporal resolution capacity than that of the FDTD methods in the nonrelativistic limit regime and carry out stability and convergence analysis for these new methods. Again, for simplicity of notations, we only present the schemes and their analysis for the 1D problem with periodic boundary conditions, i.e. (2.1). Generalization to higher dimensions is straightforward and the error estimates remain valid without modifications.

3.1 Numerical methods

First we present the Gautschi-type exponential integrator Fourier pseudospectral (Gautschi-FP) method which is based on the application of Fourier pseudospectral approach to spatial discretization followed by a Gautschi-type exponential integrator [16,21,22,24,25] to time discretization. Let

$$Y_M = \text{span}\{\phi_l(x) = \mathrm{e}^{i\mu_l(x-a)},\ -M/2 \le l \le M/2-1\}.$$

For any periodic function $v(x)$ on $[a,b]$ and vector $v \in X_M$, define $P_M : L^2(a,b) \to Y_M$ as the standard projection operator, $I_M : C(a,b) \to Y_M$ and $I_M : X_M \to Y_M$ as the trigonometric interpolation operators [46], i.e.

$$(P_M v)(x) = \sum_{l=-M/2}^{M/2-1} \widehat{v}_l \mathrm{e}^{i\mu_l(x-a)}, \quad (I_M v)(x) = \sum_{l=-M/2}^{M/2-1} \widetilde{v}_l \mathrm{e}^{i\mu_l(x-a)}, \quad a \le x \le b,$$

with $\left(l = -\frac{M}{2}, \ldots, \frac{M}{2}-1\right)$

$$\widehat{v}_l = \frac{1}{b-a}\int_a^b v(x)\mathrm{e}^{-i\mu_l(x-a)}\mathrm{d}x, \quad \widetilde{v}_l = \frac{1}{M}\sum_{j=0}^{M-1} v_j \mathrm{e}^{-i\mu_l(x_j-a)},$$

where v_j is interpreted as $v(x_j)$ for the periodic function $v(x)$. In addition, we adopt the same notation as vector case (2.9) to define the discrete l^2-norm for the periodic function $v(x)$ as $\|v\|_{l^2}^2 = h\sum_{j=0}^{M-1}|v(x_j)|^2$.

The Fourier spectral method for (2.1) is as follows:
Find $u_M(x,t) \in Y_M$, i.e.

$$u_M(x,t) = \sum_{l=-M/2}^{M/2-1} \widehat{u}_l(t) \mathrm{e}^{i\mu_l(x-a)}, \quad a \le x \le b, \quad t \ge 0, \tag{3.1}$$

such that

$$\varepsilon^2 \partial_{tt} u_M(x,t) - \Delta u_M + \frac{1}{\varepsilon^2} u_M + P_M f(u_M) = 0, \quad a \le x \le b, \quad t \ge 0. \tag{3.2}$$

Plugging (3.1) into (3.2), noticing the orthogonality of the Fourier functions, we find

$$\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} \widehat{u}_l(t) + \frac{1+\varepsilon^2\mu_l^2}{\varepsilon^2} \widehat{u}_l(t) + \widehat{f(u_M)}_l(t) = 0, \quad l = -\frac{M}{2}, \dots, \frac{M}{2} - 1, \quad t \ge 0. \tag{3.3}$$

For each fixed $l (l = -\frac{M}{2}, -\frac{M}{2}+1, \dots, \frac{M}{2}-1)$, when t is near $t = t_n (n \ge 0)$, we re-write the above ODEs as

$$\frac{\mathrm{d}^2}{\mathrm{d}w^2} \widehat{u}_l(t_n + w) + (\beta_l^n)^2 \widehat{u}_l(t_n + w) + \frac{1}{\varepsilon^2} \widehat{g}_l^n(w) = 0, \quad w \in \mathbb{R}, \tag{3.4}$$

where

$$\beta_l^n = \frac{1}{\varepsilon^2} \sqrt{1 + \varepsilon^2(\mu_l^2 + \alpha^n)}, \quad \widehat{g}_l^n(w) = \widehat{f(u_M)}_l(t_n + w) - \alpha^n \widehat{u}_l(t_n + w). \tag{3.5}$$

Here we introduce a linear stabilization term with stabilizing constant α^n satisfying $1 + \varepsilon^2\alpha^n > 0$ such that the scheme is unconditionally stable (see below for its choice). Using the variation-of-constants formula as in the Gautschi-type exponential integrator for oscillatory second-order differential equations [21,22,25,24], the general solution of the above second-order ODEs can be written as

$$\widehat{u}_l(t_n + w) = c_l^n \cos(w\beta_l^n) + d_l^n \frac{\sin(w\beta_l^n)}{\beta_l^n} - \frac{1}{\varepsilon^2 \beta_l^n} \int_0^w \widehat{g}_l^n(s) \sin(\beta_l^n(w - s)) \mathrm{d}s, \tag{3.6}$$

where c_l^n and d_l^n are two constants to be determined.

Now the key problem is how to choose two proper transmission conditions for the second-order ODEs (3.4) between different time intervals so that we can uniquely determine the two constants in (3.6). When $n = 0$, considering the solution (3.6) for $w \in [0, \tau]$, from the initial conditions in (2.1) these two conditions can be chosen naturally as

$$\widehat{u}_l(0) = \widehat{\phi}_l, \quad \frac{\mathrm{d}}{\mathrm{d}t} \widehat{u}_l(0) = \frac{1}{\varepsilon^2} \widehat{\gamma}_l. \tag{3.7}$$

Plugging (3.7) into (3.6) with $n = 0$ to determine the two constants c_l^0 and d_l^0 and then letting $w = \tau$, we get

$$\widehat{u}_l(\tau) = \widehat{\phi}_l \cos(\tau\beta_l^0) + \widehat{\gamma}_l \frac{\sin(\tau\beta_l^0)}{\varepsilon^2\beta_l^0} - \frac{1}{\varepsilon^2\beta_l^0}\int_0^\tau \widehat{g}_l^{\,0}(s)\sin\left(\beta_l^0(\tau-s)\right)\mathrm{d}s. \quad (3.8)$$

For $n > 0$, we consider the solution in (3.6) for $w \in [-\tau, \tau]$ and require the solution to be continuous at $t = t_n$ and $t = t_{n-1} = t_n - \tau$. Plugging $w = 0$ and $w = -\tau$ into (3.6) to determine the two constants c_l^n and d_l^n and then letting $w = \tau$, noticing (3.5), we have

$$\begin{aligned}\widehat{u}_l(t_{n+1}) &= -\widehat{u}_l(t_{n-1}) + 2\cos(\tau\beta_l^n)\widehat{u}_l(t_n)\\ &\quad -\frac{1}{\varepsilon^2\beta_l^n}\int_0^\tau \left[\widehat{g}_l^{\,n}(-s) + \widehat{g}_l^{\,n}(s)\right]\sin(\beta_l^n(\tau-s))\mathrm{d}s. \qquad (3.9)\end{aligned}$$

In order to design an explicit scheme, we approximate the integrals in (3.8) and (3.9) by the following quadrature

$$\int_0^\tau \widehat{g}_l^{\,0}(s)\sin(\beta_l^0(\tau-s))\mathrm{d}s \approx \widehat{g}_l^{\,0}(0)\int_0^\tau \sin\left(\beta_l^0(\tau-s)\right)\mathrm{d}s = \frac{\widehat{g}_l^{\,0}(0)}{\beta_l^0}\left[1-\cos(\tau\beta_l^0)\right],$$

$$\int_0^\tau \left[\widehat{g}_l^{\,n}(-s)+\widehat{g}_l^{\,n}(s)\right]\sin(\beta_l^n(\tau-s))\mathrm{d}s \approx \frac{2\widehat{g}_l^{\,n}(0)}{\beta_l^n}\left[1-\cos(\tau\beta_l^n)\right].$$

Denote $\widehat{(u_M^n)}_l$ and $u_M^n(x)$ be the approximations of $\widehat{u}_l(t_n)$ and $u_M(x, t_n)$, respectively. Choosing $u_M^0(x) = (P_M\phi)(x)$ and noticing (3.5), then a Gautschi-type exponential integrator Fourier spectral discretization for the KG equation (2.1) is:

$$u_M^{n+1}(x) = \sum_{l=-M/2}^{M/2-1} \widehat{(u_M^{n+1})}_l\, \mathrm{e}^{i\mu_l(x-a)}, \quad a \le x \le b, \quad n = 0, 1, \ldots, \qquad (3.10)$$

where

$$\widehat{(u_M^1)}_l = p_l^0\,\widehat{\phi}_l + q_l^0\,\widehat{\gamma}_l + r_l^0\,\widehat{(f(\phi))}_l, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2}-1, \qquad (3.11)$$

$$\widehat{(u_M^{n+1})}_l = -\widehat{(u_M^{n-1})}_l + p_l^n\,\widehat{(u_M^n)}_l + r_l^n\,\widehat{(f(u_M^n))}_l, \quad n \ge 1; \qquad (3.12)$$

with

$$p_l^0 = \cos(\tau\beta_l^0) + \frac{\alpha^0(1-\cos(\tau\beta_l^0))}{\left(\varepsilon\beta_l^0\right)^2}, \quad q_l^0 = \frac{\sin(\tau\beta_l^0)}{\varepsilon^2\beta_l^0}, \quad r_l^0 = \frac{\cos(\tau\beta_l^0)-1}{(\varepsilon\beta_l^0)^2}, \tag{3.13}$$

$$p_l^n = 2\left[\cos(\tau\beta_l^n) + \frac{\alpha^n(1-\cos(\tau\beta_l^n))}{(\varepsilon\beta_l^n)^2}\right], \quad r_l^n = \frac{2(\cos(\tau\beta_l^n)-1)}{(\varepsilon\beta_l^n)^2}, \quad n \geq 1. \tag{3.14}$$

As demonstrated in the literature [16,21,22,24,25], the above Gautschi-type exponential integrator gives exact solution for the linear second-order ODEs (3.4) and has favorable properties compared to standard time integrators for oscillatory second-order ODEs. In the next two subsections, we will demonstrate that the above discretization gives exact solution in time for the linear KG equation (2.1), i.e. $f(u) = \alpha u$, under the choice of $\alpha^n = \alpha (n \geq 0)$ in (3.4); and respectively, performs much better resolution in time than that of the FDTD methods for the nonlinear KG equation. We remark that similar techniques for time discretization have been used in discretizing wave-type equations in Zakharov system [6], Maxwell–Dirac equations [7] and Klein–Gordon–Schrödinger equations [8].

The above procedure is not suitable in practice due to the difficulty of computing the integrals in (3.11) and (3.12). We now present an efficient implementation by choosing $u_M^0(x)$ as the interpolation of $\phi(x)$ on the grids $\{x_j,\ j = 0, 1, \ldots, M\}$, i.e. $u_M^0(x) = (I_M\phi)(x)$, and approximating the integrals in (3.11) and (3.12) by a quadrature rule on the grids. Let u_j^n be the approximation of $u(x_j, t_n)$ and denote $u_j^0 = \phi(x_j)(j = 0, 1, \ldots, M)$. For $n = 0, 1, \ldots$, a Gautschi-type exponential integrator Fourier pseudospectral (Gautschi-FP) discretization for the KG equation (2.1) is

$$u_j^{n+1} = \sum_{l=-M/2}^{M/2-1} \widetilde{(u^{n+1})}_l \, e^{2ijl\pi/M}, \quad j = 0, 1, \ldots, M, \tag{3.15}$$

where,

$$\widetilde{(u^1)}_l = p_l^0\, \widetilde{\phi}_l + q_l^0\, \widetilde{\gamma}_l + r_l^0\, \widetilde{(f(\phi))}_l, \quad l = -\frac{M}{2}, \ldots, \leq \frac{M}{2} - 1,$$
$$\widetilde{(u^{n+1})}_l = -\widetilde{(u^{n-1})}_l + p_l^n\, \widetilde{(u^n)}_l + r_l^n\, \widetilde{(f(u^n))}_l, \quad n \geq 1,$$

with p_l^n, q_l^0 and r_l^n are given in (3.13) and (3.14). Based on the results in Theorem 6 (see below), in practice, we suggest that α^n is chosen as: If $f(v) = \alpha v$ is a linear function with α a constant, we choose $\alpha^n = \max\{-1/\varepsilon^2,\ \alpha\}$ for $n \geq 0$, and respectively; if $f(v)$ is a nonlinear function, we choose $\alpha^{-1} = 0$ and

$$\alpha^n = \max\left\{\alpha^{n-1}, \max_{u_j^n \neq 0,\ 0\leq j\leq M} f(u_j^n)/u_j^n\right\}, \quad n \geq 0. \tag{3.16}$$

This Gautschi-FP discretization is explicit, time symmetric and easy to extend to 2D and 3D. The memory cost is $O(M)$ and computational cost per time step is $O(M \ln M)$ via FFT.

Remark 2 Another way to approximate the integrals in (3.8) and (3.9) is to use the trapezoidal rule:

$$\int_0^\tau \widehat{g}_l^0(s)\sin(\beta_l^0(\tau-s))\mathrm{d}s \approx \frac{\tau}{2}\widehat{g}_l^0(0)\sin(\tau\beta_l^0),$$

$$\int_0^\tau \left[\widehat{g}_l^n(-s)+\widehat{g}_l^n(s)\right]\sin(\beta_l^n(\tau-s))\mathrm{d}s \approx \tau\,\widehat{g}_l^n(0)\sin(\tau\beta_l^n).$$

The rest of computations can be carried out in a similar manner.

For comparison, here we also introduce the Gautschi-type exponential integrator finite difference (Gautschi-FD) method which is based on applying centered finite difference to spatial discretization followed by a Gautschi-type integrator to time discretization. The aim is to show that the temporal resolution capacity of the Gautschi-type integrator for wave-type equation is independent of the spatial discretization that it follows [23]. Let $u_j(t)$ be the approximation of $u(x_j,t)(j=0,1,\ldots,M)$. Applying a centered finite difference to the spatial derivative in (2.1a), we get

$$\varepsilon^2\frac{\mathrm{d}^2}{\mathrm{d}t^2}u_j(t)-\delta_x^2u_j(t)+\frac{1}{\varepsilon^2}u_j(t)+f(u_j(t))=0,\quad 0\le j\le M-1, \tag{3.17}$$

with $u_0(t)=u_M(t)$ and $u_{-1}(t)=u_{M-1}(t)$. Let $U(t)=(u_0(t),u_1(t),\ldots,u_{M-1}(t))^T$ and $F(U(t))=(f(u_0(t)),f(u_1(t)),\ldots,f(u_{M-1}(t)))^T$, then the above ODEs can be rewritten as

$$\varepsilon^2U''(t)+A\,U(t)+F(U(t))=0,\quad t\ge 0, \tag{3.18}$$

where A is an $M\times M$ matrix independent of t. Since A is symmetric, it is normal, i.e. there exists an orthogonal matrix P and a diagonal matrix Λ such that

$$A=P^{-1}\,\Lambda\,P.$$

Let $V(t)=P\;U(t)$ and multiply P to both sides of (3.18), we get

$$\varepsilon^2V''(t)+\Lambda V(t)+P\;F(U(t))=0,\quad t\ge 0. \tag{3.19}$$

The above second-order ODEs are similar as (3.4) and we apply the Gautschi-type exponential integrator to discretize it which immediately gives a discretization of (3.17). We omit the details of the derivation for brevity and conclude that the scheme is the same as (3.15), with μ_l in (3.5) replaced by λ_l defined in (2.19).

3.2 Stability and convergence analysis in linear case

In this subsection, we assume that $f(u)$ is a linear function, i.e. $f(u) = \alpha u$ with α being a constant satisfying $\alpha > -\varepsilon^{-2}$. In this case, the solution of (2.1) is

$$u(x,t) = \sum_{l=-\infty}^{\infty} \left[\widehat{\phi}_l \cos(t\beta_l) + \widehat{\gamma}_l \frac{\sin(t\beta_l)}{\varepsilon^2 \beta_l} \right] e^{i\mu_l(x-a)}, \quad a \le x \le b,\ t \ge 0, \tag{3.20}$$

where

$$\beta_l = \frac{1}{\varepsilon^2}\sqrt{1 + \varepsilon^2(\mu_l^2 + \alpha)}, \quad l = 0, \pm 1, \ldots. \tag{3.21}$$

Again, by using the standard von Neumann analysis [45], we have the following stability results for Gautschi-FP and Gautschi-FD:

Theorem 6 *If α^n in (3.5) is chosen such that $\alpha^n \ge \alpha$ for $n \ge 0$, then both Gautschi-FP and Gautschi-FD are unconditionally stable for any $\tau > 0$, $h > 0$ and $\varepsilon > 0$.*

Proof Similar to the proof of Theorem 1, noticing (3.14) and (3.15), we have the same characteristic equation (2.16) for Gautschi-FP with

$$\begin{aligned} \theta_l &= \cos(\tau\beta_l^n) + \frac{(\alpha^n - \alpha)(1 - \cos(\tau\beta_l^n))}{(\varepsilon\beta_l^n)^2} \\ &= \cos^2\left(\frac{\tau\beta_l^n}{2}\right) + \left[\frac{2(\alpha^n - \alpha)}{\varepsilon^{-2} + \mu_l^2 + \alpha^n} - 1\right]\sin^2\left(\frac{\tau\beta_l^n}{2}\right), \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1. \end{aligned}$$

Since $\alpha^n \ge \alpha > -\varepsilon^{-2} (n \ge 0)$, we have

$$0 \le \frac{2(\alpha^n - \alpha)}{\varepsilon^{-2} + \mu_l^2 + \alpha^n} \le 2 \implies |\theta_l| \le 1, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1, \tag{3.22}$$

which immediately leads to the unconditional stability of the Gautschi-FP. For Gautschi-FD, we only need to replace μ_l in (3.22) by λ_l defined in (2.19) and the stability claim follows immediately. □

Let $u_I(x,t)$ be the solution of the following problem

$$\varepsilon^2 \partial_{tt} u_I(x,t) - \partial_{xx} u_I + \frac{1}{\varepsilon^2} u_I + \alpha u_I = 0, \quad a < x < b, \quad t > 0, \tag{3.23a}$$

$$u_I(a,t) = u_I(b,t), \quad \partial_x u_I(a,t) = \partial_x u_I(b,t), \quad t \ge 0, \tag{3.23b}$$

$$u_I(x,0) = (I_M\phi)(x), \quad \partial_t u_I(x,0) = \frac{1}{\varepsilon^2}(I_M\gamma)(x), \quad a \le x \le b. \tag{3.23c}$$

It is easy to see that the solution of the above problem is

$$u_I(x,t) = \sum_{l=-M/2}^{M/2-1} \left[\cos(t\beta_l)\widetilde{\phi}_l + \widetilde{\gamma}_l \frac{\sin(t\beta_l)}{\varepsilon^2 \beta_l} \right] \mathrm{e}^{i\mu_l(x-a)}, \quad a \le x \le b,\ t \ge 0. \quad (3.24)$$

Denote

$$e_j^n = u(x_j, t_n) - u_j^n, \quad j = 0, 1, \ldots, M, \quad n \ge 0,$$
$$e^n(x) := u(x, t_n) - (I_M u^n)(x), \quad a \le x \le b, \quad n \ge 0.$$

For Gautschi-FP, we have the following error estimates (see detailed proof in Appendix III below):

Theorem 7 *Let u_j^n be the solution of Gautschi-FP (3.15) with $\alpha^n = \alpha$ in (3.5) for $n \ge 0$. Then we have*

$$u_j^n = u_I(x_j, t_n), \quad j = 0, 1, \ldots, M, \quad n \ge 0. \quad (3.25)$$

In addition, if $\phi,\ \gamma \in H_p^m := \{v \in H^m(a,b) \mid v^{(l)}(a) = v^{(l)}(b), 0 \le l \le m-1\}$ with $m \ge 2$, when $\alpha \ge 0$ for any $\varepsilon > 0$ or when $\alpha < 0$ for $0 < \varepsilon \le \varepsilon_0 := \frac{1}{\sqrt{2|\alpha|}}$, we have the following error estimates

$$\|e^n(x)\|_{L^2} \lesssim h^m, \quad \|\nabla e^n(x)\|_{L^2} \lesssim h^{m-1}, \quad n \ge 0. \quad (3.26)$$

Thus if the functions ϕ and γ are periodic and smooth, for the linear KG equation, the Gautschi-FP converges exponentially fast in space with no error in time discretization.

Also, we have error estimates for Gautschi-FD in linear case (see detailed proof in Appendix IV below):

Theorem 8 *Let u_j^n be the solution of Gautschi-FD with $\alpha^n = \alpha$ for $n \ge 0$. If $\phi,\ \gamma \in W_p^{4,\infty}(\Omega)$, when $\alpha \ge 0$ for any $\varepsilon > 0$ or when $\alpha < 0$ for $0 < \varepsilon \le \varepsilon_0 := \frac{1}{\sqrt{2|\alpha|}}$, we have*

$$\|e^n\|_{l^2} \lesssim h^2, \quad 0 \le n \le \frac{T}{\tau}. \quad (3.27)$$

Based on Theorems 7 and 8, both Gautschi-FP and Gautschi-FD introduce no error from time discretization for the linear KG equation, and share the same temporal resolution in the nonrelativistic limit regime. In fact, for a given accuracy $\delta > 0$, for the linear KG equation the ε-scalability of the two methods is:

$$\tau = O(1), \quad h \le O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1, \quad (3.28)$$

i.e. both mesh size h and time step τ can be chosen independently of the small parameter ε.

3.3 Convergence analysis in the nonlinear case

In order to obtain an error estimate for Gautschi-FP (3.10) with (3.16), let $0 < T < T^*$ with T^* the maximum existence time of the solution, motivated by the results in [32,33], we assume that there exists an integer $m_0 \geq 2$ such that

$$\text{(C)} \quad u \in C^2\left([0,T]; H^1\right) \cap C^1\left([0,T]; W^{1,4}\right) \cap C\left([0,T]; L^\infty \cap H_p^{m_0}\right),$$
$$\|\partial_t u(x,t)\|_{L^\infty([0,T];W^{1,4})} \lesssim \frac{1}{\varepsilon^2}, \quad \|\partial_{tt} u(x,t)\|_{L^\infty([0,T];H^1)} \lesssim \frac{1}{\varepsilon^4},$$
$$\|u(x,t)\|_{L^\infty([0,T];L^\infty\cap H_p^{m_0})} \lesssim 1.$$

Under the above assumption (C) and assume $f \in C^3(\mathbb{R})$, we have

$$M_1 := \max_{0\leq t\leq T} \|u(x,t)\|_{L^\infty} \lesssim 1, \quad M_2 := \max_{|v|\leq 1+M_1} \sum_{l=1}^{3} |f^{(l)}(v)| \lesssim 1, \tag{3.29}$$

$$M_3 := \max\left\{0, \sup_{0\neq v,\, |v|\leq 1+M_1} f(v)/v\right\} \leq M_2 \lesssim 1. \tag{3.30}$$

Assuming

$$\tau \leq \frac{\pi\varepsilon^2 h}{3\sqrt{h^2+\varepsilon^2(\pi^2+M_3h^2)}}, \tag{3.31}$$

we have (see detailed proof in Appendix V below):

Theorem 9 *Let $u_M^n(x)$ be the approximation obtained from the Gautschi-FP method* (3.10) *with* (3.16)*. Assume $\tau \lesssim \varepsilon^2\sqrt{C_d(h)}$ and $f(\cdot) \in C^3(\mathbb{R})$, under the assumption (C), there exist $h_0 > 0$ and $\tau_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \leq 1$, when $0 < h \leq h_0$ and $0 < \tau \leq \tau_0$ and under the condition* (3.31)*, we have the following error estimate*

$$\|u(x,t_n) - u_M^n(x)\|_{L^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^{m_0}, \quad \|u_M^n(x)\|_{L^\infty} \leq 1 + M_1, \tag{3.32a}$$

$$\|\nabla[u(x,t_n) - u_M^n(x)]\|_{L^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^{m_0-1}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{3.32b}$$

Similar to the proof for the above theorem, for Gautschi-FD with (3.16), assume that $f(\cdot) \in C^2(\mathbb{R}), u \in C^2([0,T];L^\infty) \cap C^1([0,T];L^\infty) \cap C([0,T];W_p^{4,\infty})$, $\|u(x,t)\|_{L^\infty(\Omega_T)} + \|\partial_{xxxx} u(x,t)\|_{L^\infty(\Omega_T)} \lesssim 1$ and

$$\|\partial_t u(x,t)\|_{L^\infty(\Omega_T)} \lesssim \frac{1}{\varepsilon^2}, \quad \|\partial_{tt} u(x,t)\|_{L^\infty(\Omega_T)} \lesssim \frac{1}{\varepsilon^4},$$

then we can prove the following error estimate for Gautschi-FD with (3.16):

Theorem 10 *Let u_j^n be the approximation obtained from Gautschi-FD with (3.16). Assume $\tau \lesssim \varepsilon^2 h$, under the above assumptions on the exact solution u and the nonlinear function f, there exist $h_0 > 0$ and $\tau_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < h \le h_0$ and $0 < \tau \le \tau_0$ and under the condition (3.31), we have*

$$\|e^n\|_{l^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^2, \quad \|u^n\|_{L^\infty} \le 1 + M_1, \quad 0 \le n \le \frac{T}{\tau}, \tag{3.33}$$

where

$$e^n = (e_0^n, e_1^n, \ldots, e_M^n)^T, \quad \textit{with} \quad e_j^n = u(x_j, t_n) - u_j^n, \quad 0 \le j \le M, \quad n \ge 0.$$

Proof Follow the analogous proofs to Theorems 9 and 8 and we omit the details here for brevity. □

Based on Theorems 9 and 10, for the nonlinear KG equation in the nonrelativistic limit regime, the ε-scalability of Gautschi-FP and Gautschi-FD is:

$$\tau = O\left(\sqrt{\delta}\varepsilon^2\right) = O(\varepsilon^2), \quad h \le O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1. \tag{3.34}$$

Remark 3 When the stabilization factor in (3.5) is chosen as $\alpha^n \equiv 0$ for all $n \ge 0$, the error estimates in Theorems 9 and 10 are still valid.

4 Numerical results

In this section, we report numerical results to support our error estimates and demonstrate the superiority of Gautschi-type integrator over finite difference in time resolution when $0 < \varepsilon \ll 1$. In order to do so, in the KG equation (2.1), we choose

$$f(u) = \lambda u^{p+1}; \quad \phi(x) = \frac{2}{\mathrm{e}^{x^2} + \mathrm{e}^{-x^2}}, \quad \gamma(x) = 0, \quad x \in \mathbb{R}. \tag{4.1}$$

The computational interval $[a, b]$ is chosen large enough such that the periodic boundary conditions do not introduce a significant aliasing error relative to the problem in the whole space. Let $u(x, t)$ be the 'exact' solution which is obtained numerically by using Gautschi-FP with very fine mesh size and small time step, e.g. $h = 1/1{,}024$ and $\tau = 1\text{E}{-8}$. In order to quantify the convergence, we define three error functions, l^2-error, l^∞-error and discrete H^1-error as

$$e_{l^2} = \|u(\cdot, t_n) - u^n\|_{l^2}, \quad e_{l^\infty} = \max_j |u(x_j, t_n) - u_j^n|,$$
$$e_{H^1} = \sqrt{\|u(\cdot, t_n) - u^n\|_{l^2}^2 + \|\delta_x^+(u(\cdot, t_n) - u^n)\|_{l^2}^2}.$$

Table 1 Temporal discretization errors of Impt-EC-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^3)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows)

ε-Scalability	τ =1.00E–3	τ =5.00E–4	τ =2.50E–4	τ =1.25E–4	τ =6.25E–5
$\varepsilon = 0.1,\ \tau$	4.6484E–2	1.1063E–2	2.7344E–3	6.8472E–4	1.7533E–4
$\varepsilon/2,\ \tau/2^3$	4.9171E–2	1.2912E–2	3.2712E–3	8.2486E–4	2.1197E–4
$\varepsilon/4,\ \tau/4^3$	4.6831E–2	1.1162E–2	2.7597E–3	6.9083E–4	1.7681E–4
$\varepsilon/8,\ \tau/8^3$	3.6900E–2	9.6406E–3	2.4426E–3	6.1784E–4	1.6129E–4
$\varepsilon = 0.1,\ \tau$	7.5093E–2	1.8650E–2	4.6877E–3	1.2087E–3	2.8179E–4
$\varepsilon/2,\ \tau/2^3$	9.2221E–2	2.3739E–2	6.0030E–3	1.5347E–3	3.8945E–4
$\varepsilon/4,\ \tau/4^3$	7.0780E–2	1.7431E–2	4.3724E–3	1.1292E–3	2.9825E–4
$\varepsilon/8,\ \tau/8^3$	7.7202E–2	1.9840E–2	5.0233E–3	1.2937E–3	3.1687E–4
$\varepsilon = 0.1,\ \tau$	2.9725E–2	7.7927E–3	1.8177E–3	4.5897E–4	1.2252E–4
$\varepsilon/2,\ \tau/2^3$	4.3783E–2	1.1543E–2	2.9273E–3	7.3938E–4	1.9031E–4
$\varepsilon/4,\ \tau/4^3$	2.8754E–2	6.9944E–3	1.7321E–3	4.2850E–4	1.1127E–4
$\varepsilon/8,\ \tau/8^3$	2.9213E–2	7.9193E–3	2.0227E–3	5.1313E–4	1.3350E–4

Case I. A nonlinear case, where we choose $\lambda = 4$ and $p = 2$ in (4.1) and solve the KG equation (2.1) on the interval $[-8, 8]$. In order to study the temporal resolution or ε-scalability in time of different methods, we choose a very small mesh size $h = 1/128$ such that the discretization error in space is negligible. Tables 1 and 2 tabulate l^2-error, H^1-error and l^∞-error at time $t = 0.4$ of Impt-EC-FD and SImpt-FD, respectively, for various time steps τ and parameter values ε under ε-scalability $\tau = O(\varepsilon^3)$. Tables 3 and 4 show similar results for Gautschi-FP and Gautschi-FD, under ε-scalability $\tau = O(\varepsilon^2)$. Similarly, in order to compare errors of spatial discretization, we always choose very fine time step τ such that time discretization error is negligible. Table 5 lists l^2-errors at time $t = 0.4$ of Impt-EC-FD, SImpt-EC-FD, Gautschi-FD and Gautschi-FP with different ε and τ satisfying the required ε-scalability. We also carried out numerical experiments for Impt-EC-FD and Exmpt-FD, where the results are similar to those of Impt-EC-FD and SImpt-FD, and thus we omit them for brevity.

Case II. A linear case, where we choose $\lambda = 4$ and $p = 0$ in (4.1) and solve the KG equation (2.1) on the interval $[-16, 16]$. Here we only present the results of Gautschi-FP and Gautschi-FD to verify that there is no time discretization error of Gautschi-type integrator for the linear KG equation. Table 6 lists the l^2-error of Gautschi-FP and Gautschi-FD at time $t = 1$ for different τ, h and ε under the ε-scalability $\tau = O(1)$ and $h = O(1)$. Similar convergence patterns of the discrete H^1-error and l^∞-error were also observed and they are omitted here for simplicity. In addition, the results for FDTD methods are quite similar to those in nonlinear case and thus are omitted here too for brevity.

From Tables 1, 2, 3, 4, 5 and 6 and our extensive numerical results not shown here for brevity, we can draw the following conclusions:

Table 2 Temporal discretization errors of SImpt-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^3)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows)

ε-Scalability	τ =1.00E–3	τ =5.00E–4	τ =2.50E–4	τ =1.25E–4	τ =6.25E–5
$\varepsilon = 0.1,\ \tau$	4.4395E–2	1.0598E–2	2.6213E–3	6.5674E–4	1.6847E–4
$\varepsilon/2,\ \tau/2^3$	4.8329E–2	1.2678E–2	3.2113E–3	8.0980E–4	2.0824E–4
$\varepsilon/4,\ \tau/4^3$	4.6690E–2	1.1131E–2	2.7521E–3	6.8896E–4	1.7635E–4
$\varepsilon/8,\ \tau/8^3$	3.6864E–2	9.6301E–3	2.4399E–3	6.1716E–4	1.6113E–4
$\varepsilon = 0.1,\ \tau$	7.2046E–2	1.7868E–2	4.4913E–3	1.1605E–3	2.9095E–4
$\varepsilon/2,\ \tau/2^3$	9.0650E–2	2.3316E–2	5.8955E–3	1.5080E–3	3.8325E–4
$\varepsilon/4,\ \tau/4^3$	7.0580E–2	1.7381E–2	4.3599E–3	1.1261E–3	2.9257E–4
$\varepsilon/8,\ \tau/8^3$	7.7123E–2	1.9818E–2	5.0178E–3	1.2923E–3	3.1356E–4
$\varepsilon = 0.1,\ \tau$	2.8663E–2	7.0385E–3	1.7558E–3	4.4435E–4	1.1932E–4
$\varepsilon/2,\ \tau/2^3$	4.2804E–2	1.1276E–2	2.8593E–3	7.2231E–4	1.8604E–4
$\varepsilon/4,\ \tau/4^3$	2.8674E–2	6.9762E–3	1.7277E–3	4.2753E–4	1.1111E–4
$\varepsilon/8,\ \tau/8^3$	2.9168E–2	7.9066E–3	2.0194E–3	5.1231E–4	1.3329E–4

Table 3 Temporal discretization errors of Gautschi-FP at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^2)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows)

ε-Scalability	τ =5.00E–3	τ =2.50E–3	τ =1.25E–3	τ =6.25E–4	τ =3.125E–4
$\varepsilon = 0.1,\ \tau$	2.4902E–3	6.1124E–4	1.5208E–4	3.7957E–5	9.4697E–6
$\varepsilon/2,\ \tau/2^2$	3.1009E–3	7.6212E–4	1.8973E–4	4.7384E–5	1.1845E–5
$\varepsilon/4,\ \tau/4^2$	2.5929E–3	6.3666E–4	1.5846E–4	3.9564E–5	9.8826E–6
$\varepsilon/8,\ \tau/8^2$	2.5965E–3	6.3757E–4	1.5862E–4	3.9563E–5	9.8072E–6
$\varepsilon = 0.1,\ \tau$	6.0409E–3	1.4857E–3	3.6976E–4	9.2230E–5	2.2948E–5
$\varepsilon/2,\ \tau/2^2$	8.6467E–3	2.1232E–3	5.2845E–4	1.3197E–4	3.2989E–5
$\varepsilon/4,\ \tau/4^2$	6.3003E–3	1.5450E–3	3.8453E–4	9.6000E–5	2.3974E–5
$\varepsilon/8,\ \tau/8^2$	7.9670E–3	1.9557E–3	4.8650E–4	1.2126E–4	3.0079E–5
$\varepsilon = 0.1,\ \tau$	1.9268E–3	4.7365E–4	1.1786E–4	2.9447E–5	7.3746E–6
$\varepsilon/2,\ \tau/2^2$	2.4770E–3	6.0895E–4	1.5161E–4	3.7863E–5	9.4650E–6
$\varepsilon/4,\ \tau/4^2$	1.9261E–3	4.7358E–4	1.1797E–4	2.9445E–5	7.3572E–6
$\varepsilon/8,\ \tau/8^2$	1.9103E–3	4.6947E–4	1.1682E–4	2.9120E–5	7.2235E–6

(i). In the $O(1)$-speed of light regime, i.e. $0 < \varepsilon = O(1)$ fixed, the FDTD methods and Gautschi-FD are of second-order accuracy in both time and space (cf. Tables 1, 2, 4, 5); where Gautschi-FP is second-order and spectral-order accurate in time and space, respectively (cf. Tables 3, 5). In addition, there is no time discretization error of Gautschi-FP and Gautschi-FD for the linear KG equation (cf. Table 6). Therefore, in this regime all the methods we have considered are compatible in time discretization while Gautschi-FP is of higher

Table 4 Temporal discretization errors of Gautschi-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^2)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows)

ε-Scalability	τ =5.00E–3	τ =2.50E–3	τ =1.25E–3	τ =6.25E–4	τ =3.125E–4
$\varepsilon = 0.1,\ \tau$	2.4910E–3	6.1204E–4	1.5295E–4	3.9100E–5	9.9575E–6
$\varepsilon/2,\ \tau/2^2$	3.1013E–3	7.6248E–4	1.9017E–4	4.8121E–5	1.2658E–5
$\varepsilon/4,\ \tau/4^2$	2.5937E–3	6.3748E–4	1.5836E–4	4.0818E–5	1.0343E–6
$\varepsilon/8,\ \tau/8^2$	2.6106E–3	6.3814E–4	1.5927E–4	4.0565E–5	9.9140E–6
$\varepsilon = 0.1,\ \tau$	6.0467E–3	1.4916E–3	3.7655E–4	9.9249E–5	2.3124E–5
$\varepsilon/2,\ \tau/2^2$	8.6502E–3	2.1268E–3	5.3291E–4	1.3656E–4	3.6826E–5
$\varepsilon/4,\ \tau/4^2$	6.3067E–3	1.5698E–3	3.9225E–4	1.0798E–4	2.4152E–5
$\varepsilon/8,\ \tau/8^2$	7.8831E–3	1.9601E–3	4.9192E–4	1.2936E–4	3.7448E–5
$\varepsilon = 0.1,\ \tau$	1.9254E–3	4.7230E–4	1.1654E–4	2.8122E–5	7.5647E–6
$\varepsilon/2,\ \tau/2^2$	2.4755E–3	6.0746E–4	1.5013E–4	3.6437E–5	8.8545E–6
$\varepsilon/4,\ \tau/4^2$	1.9247E–3	4.7285E–4	1.1662E–4	2.9170E–5	7.6260E–6
$\varepsilon/8,\ \tau/8^2$	1.9340E–3	4.6890E–4	1.1568E–4	2.7885E–5	7.5793E–6

Table 5 Spatial discretization error e_{l^2} of Impt-EC-FD and SImpt-FD (under ε-scalability $\tau = O(\varepsilon^3)$) and Gautschi-FD and Gautschi-FP (under ε-scalability $\tau = O(\varepsilon^2)$) at time $t = 0.4$ in nonlinear case with $\varepsilon_0 = 0.1$ and τ_0=2E-5 for different mesh sizes h

	$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$
Impt-EC-FD				
$\varepsilon_0,\ \tau_0$	2.0671E–2	5.5497E–3	1.4075E–3	3.5551E–4
$\varepsilon_0/2,\ \tau_0/2^3$	2.2900E–2	6.2179E–3	1.5834E–3	4.0110E–4
$\varepsilon_0/4,\ \tau_0/4^3$	2.2881E–2	6.2815E–3	1.6021E–3	4.0398E–4
SImpt-FD				
$\varepsilon_0,\ \tau_0$	2.0671E–2	5.5497E–3	1.4075E–3	3.5541E–4
$\varepsilon_0/2,\ \tau_0/2^3$	2.2900E–2	6.2178E–3	1.5833E–3	4.0101E–4
$\varepsilon_0/4,\ \tau_0/4^3$	2.2881E–2	6.2815E-3	1.6021E–3	4.0398E–4
Gautschi-FD				
$\varepsilon_0,\ \tau_0$	2.0668E–2	5.5462E–3	1.4041E–3	3.5182E–4
$\varepsilon_0/2,\ \tau_0/2^2$	2.2894E–2	6.2129E–3	1.5784E–3	3.9568E–4
$\varepsilon_0/4,\ \tau_0/4^2$	2.2878E–2	6.2790E–3	1.5996E–3	4.0120E–4
	$h = 1$	$h = 1/2$	$h = 1/4$	$h = 1/8$
Gautschi-FP				
$\varepsilon_0,\ \tau_0$	1.1873E–1	3.9320E–3	3.1799E–5	1.0722E–7
$\varepsilon_0/2,\ \tau_0/2^2$	8.3243E–2	3.2486E–3	3.3677E–5	7.6844E–8
$\varepsilon_0/4,\ \tau_0/4^2$	1.1899E–1	3.9849E–3	2.8723E–5	8.4444E–8

Table 6 Temporal and spatial discretization error e_{l^2} of Gautschi-FP and Gautschi-FD in linear case at time $t = 1$ with $\tau_0 = 0.25$ and $h_0 = 0.5$ for different τ, h and ε

		$\varepsilon = 0.02$	$\varepsilon = 0.002$	$\varepsilon = 0.0002$	$\varepsilon = 0.00002$
Gautschi-FP					
τ_0	$h_0/8$	1.1239E–15	9.7781E–16	1.9602E–15	1.6371E–15
$\tau_0/2$	$h_0/8$	1.2503E–15	1.6460E–15	1.3867E–15	1.6133E–15
$\tau_0/4$	$h_0/8$	2.8930E–15	2.2077E–15	2.5100E–15	2.4671E–15
τ_0	h_0	3.9029E–3	5.5134E–3	4.1445E–3	5.5276E–3
τ_0	$h_0/2$	1.1041E–5	1.2214E–5	8.6830E–6	1.2093E–5
τ_0	$h_0/4$	4.1894E–10	5.2825E–10	5.2296E–10	4.8898E–10
Gautschi-FD					
τ_0	$h_0/32$	2.2447E–4	2.2633E–4	2.2723E–4	2.2641E–4
$\tau_0/2$	$h_0/32$	2.2447E–4	2.2633E–4	2.2723E–4	2.2641E–4
$\tau_0/4$	$h_0/32$	2.2447E–4	2.2633E–4	2.2723E–4	2.2641E–4
τ_0	$h_0/4$	1.3608E–2	1.3703E–2	1.3765E–2	1.3708E–2
τ_0	$h_0/8$	3.5636E–3	3.5923E–3	3.6069E–3	3.5934E–3
τ_0	$h_0/16$	8.9699E–4	9.0441E–4	9.0802E–4	9.0468E–4

accuracy in space than the rest. Indeed, generally Gautschi-FP performs much better in time discretization than the rest under the same time step and mesh size.

(ii). In the nonrelativistic limit regime, i.e. $0 < \varepsilon \ll 1$, for FDTD methods the 'correct' ε-scalability is $\tau = O(\varepsilon^3)$ and $h = O(1)$ which confirms our analytical results (2.30); and for Gautschi-FP and Gautschi-FD methods, the 'correct' ε-scalability is $\tau = O(1)$ and $h = O(1)$ for the linear KG equation which verifies our analytical results (3.28), and respectively, $\tau = O(\varepsilon^2)$ and $h = O(1)$ for the nonlinear KG equation which again confirms our analytical results (3.34).

In view of both temporal and spatial resolution capacities, we conclude that Gautschi-FP is the best candidate for discretizing the KG equation, especially in the nonrelativistic limit regime.

5 Conclusion

Two classes of numerical methods with different time integrations were analyzed rigorously and compared numerically for solving the KG equation in the nonrelativistic limit regime, i.e. if $0 < \varepsilon \ll 1$ or the speed of light goes to infinity. The first class are the standard second-order FDTD methods. For FDTD schemes, including energy conservative/non-conservative and implicit/semi-implicit/explicit ones, error estimates were rigorously carried out, which showed that their ε-scalability is $\tau = O(\varepsilon^3)$ with ε-independent h. The second class are based on applying the Gautschi-type exponential integrator for time discretization, which is combined with either Fourier pseudospectral (Gautschi-FP) or finite difference (Gautschi-FD) discretization in space. For the linear KG equation, the Gautschi-type time integration does not introduce error in time discretization. In addition, our rigorous error estimates suggest that the ε-scalability of Gautschi-FP and Gautschi-FD is improved to $\tau = O(1)$ and $\tau = O(\varepsilon^2)$ for the linear and nonlinear KG equations, respectively. Comparison between Gautschi-FP and Gautschi-FD also indicated that this temporal resolution competence of Gautschi-type methods is independent of the spatial discretization that it combines with. In summary, Gautschi-FP performs the best among all the methods discussed here in both nonrelativistic limit regime and $O(1)$-speed of light regime.

Acknowledgments The authors acknowledge the support from Ministry of Education of Singapore grant R-146-000-120-112 and thank the anonymous referees for their valuable comments and suggestions in improving this paper. This work was partially done while the first author was visiting Department of Mathematics at The Hong Kong University of Science and Technology in 2011 and he would like to thank Professor Yang Xiang for his warm hospitality during the visit.

Appendix I: Proof of Theorem 2 for Impt-EC-FD (2.2)

In order to prove Theorem 2, we need the following lemmas:

Lemma 2 *For any $v^n \in X_M (n \geq 0)$, the following equalities hold*

$$-h\sum_{j=0}^{M-1} v_j^n \delta_x^2 v_j^n = h\sum_{j=0}^{M-1}\left|\delta_x^+ v_j^n\right|^2 = \|\delta_x^+ v^n\|_{l^2}^2, \tag{I.1}$$

$$h\sum_{j=0}^{M-1} v_j^n v_j^{n+1} = \frac{1}{2}\|v^n\|_{l^2}^2 + \frac{1}{2}\left\|v^{n+1}\right\|_{l^2}^2 - \frac{\tau^2}{2}\|\delta_t^+ v^n\|_{l^2}^2, \tag{I.2}$$

$$\begin{aligned} h\sum_{j=0}^{M-1}\left(\delta_x^+ v_j^{n+1}\right)\left(\delta_x^+ v_j^n\right) &= \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^n\right)^2 + \left(v_{j+1}^n - v_j^{n+1}\right)^2\right] \\ &\quad - \frac{\tau^2}{h^2}\|\delta_t^+ v^n\|_{l^2}^2 \qquad n = 0, 1, \ldots. \end{aligned} \tag{I.3}$$

Proof The equality (I.1) comes from the standard summation by parts formula (see, e.g. [30]) and (I.2) comes from

$$\begin{aligned} v_j^n v_j^{n+1} &= \frac{1}{2}\left[\left(v_j^{n+1}\right)^2 + (v_j^n)^2 - \left(v_j^{n+1} - v_j^n\right)^2\right] \\ &= \frac{1}{2}\left[\left(v_j^{n+1}\right)^2 + \left(v_j^n\right)^2 - \tau^2(\delta_t^+ v_j^n)^2\right]. \end{aligned}$$

From (I.2) and a straightforward computation, we get

$$h\sum_{j=0}^{M-1}\left(\delta_x^+ v_j^{n+1}\right)\left(\delta_x^+ v_j^n\right) = \frac{1}{2}\left\|\delta_x^+ v^{n+1}\right\|_{l^2}^2 + \frac{1}{2}\|\delta_x^+ v^n\|_{l^2}^2 - \frac{\tau^2}{2}\|\delta_t^+ \delta_x^+ v^n\|_{l^2}^2$$

$$
\begin{aligned}
&= \frac{h}{2}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ v_j^{n+1}\right)^2 + \left(\delta_x^+ v_j^n\right)^2\right] - \frac{\tau^2}{2h}\sum_{j=0}^{M-1}\left(\delta_t^+ v_{j+1}^n - \delta_t^+ v_j^n\right)^2 \\
&= \frac{\tau^2}{h}\sum_{j=0}^{M-1}\left(\delta_t^+ v_{j+1}^n\right)\left(\delta_t^+ v_j^n\right) + \frac{h}{2}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ v_j^{n+1}\right)^2 + \left(\delta_x^+ v_j^n\right)^2\right] - \frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2 \\
&= \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^{n+1}\right)^2 + \left(v_{j+1}^n - v_j^n\right)^2 + 2\left(v_{j+1}^{n+1} - v_{j+1}^n\right)\left(v_j^{n+1} - v_j^n\right)\right] \\
&\quad - \frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2 \\
&= \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^n\right)^2 + \left(v_{j+1}^n - v_j^{n+1}\right)^2\right] - \frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2,
\end{aligned}
$$

which immediately implies (I.3). □

Lemma 3 *Denote the local truncation error for Impt-EC-FD (2.2) as*

$$
\begin{aligned}
\xi_j^0 &:= \delta_t^+ u(x_j, 0) - \frac{1}{\varepsilon^2}\gamma(x_j) - \frac{\tau}{2\varepsilon^2}\left[\delta_x^2\phi(x_j) - \frac{1}{\varepsilon^2}\phi(x_j) - f(\phi(x_j))\right], \\
\xi_j^n &:= \varepsilon^2\delta_t^2(u(x_j, t_n)) - \frac{1}{2}\left[\delta_x^2(u(x_j, t_{n+1})) + \delta_x^2(u(x_j, t_{n-1}))\right] \\
&\quad + \frac{1}{2\varepsilon^2}[u(x_j, t_{n+1}) + u(x_j, t_{n-1})] + G(u(x_j, t_{n+1}), u(x_j, t_{n-1})),\ \ n \ge 1.
\end{aligned}
\tag{I.4}
$$

Assume $\tau \lesssim h$ and under the assumptions (A) and (B2), we have

$$
\|\xi^n\|_{l^2} + \left\|\delta_x^+\xi^n\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}; \quad \left\|\delta_x^2\xi^0\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}. \tag{I.5}
$$

Proof Taking Taylor's expansion in the local truncation error (I.4), noticing (2.6), (2.1), under $\tau \lesssim h$ and using the assumptions (A) and (B2), with the help of the triangle inequality and Cauchy–Schwartz inequality, for $j = 0, 1, \ldots, M-1$, we have

$$
\left|\xi_j^0\right| \le \frac{\tau^2}{6}\|\partial_{ttt}u\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\|\phi'''\|_{L^\infty(\Omega)} \lesssim \frac{\tau^2}{\varepsilon^6} + \frac{h\tau}{\varepsilon^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \tag{I.6}
$$

$$
\left|\delta_x^+\xi_j^0\right| \le \frac{\tau^2}{6}\|\partial_{tttx}u\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\|\phi''''\|_{L^\infty(\Omega)} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \tag{I.7}
$$

$$
\left|\delta_x^2\xi_j^0\right| \le \frac{\tau^2}{6}\|\partial_{tttxx}u\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\|\phi'''''\|_{L^\infty(\Omega)} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \tag{I.8}
$$

$$
\left|\xi_j^n\right| \le \frac{\varepsilon^2\tau^2}{12}\|\partial_{tttt}u\|_{L^\infty(\Omega_T)} + \frac{\tau^2}{2}\|\partial_{ttxx}u\|_{L^\infty(\Omega_T)} + \frac{h^2}{12}\|\partial_{xxxx}u\|_{L^\infty(\Omega_T)}
$$

$$+\tau^2\left[\|f'\|_{L^\infty(\mathbb{R})}\|\partial_{tt}u\|_{L^\infty(\Omega_T)}+\|f''\|_{L^\infty(\mathbb{R})}\|\partial_t u\|^2_{L^\infty(\Omega_T)}+\frac{1}{2\varepsilon^2}\|\partial_{tt}u\|_{L^\infty(\Omega_T)}\right]$$

$$\lesssim h^2+\frac{\tau^2}{\varepsilon^6}+\frac{\tau^2}{\varepsilon^4}\lesssim h^2+\frac{\tau^2}{\varepsilon^6},\quad 1\le n\le\frac{T}{\tau}, \tag{I.9}$$

$$\left|\delta_x^+\xi_j^n\right|\le\frac{\varepsilon^2\tau^2}{12}\|\partial_{ttttx}u\|_{L^\infty(\Omega_T)}+\frac{\tau^2}{2}\|\partial_{ttxxx}u\|_{L^\infty(\Omega_T)}+\frac{h^2}{12}\|\partial_{xxxxx}u\|_{L^\infty(\Omega_T)}$$
$$+\tau^2\Big[\|f''\|_{L^\infty(\mathbb{R})}\|\partial_{tt}u\|_{L^\infty(\Omega_T)}\|\partial_x u\|_{L^\infty(\Omega_T)}+\|f'\|_{L^\infty(\mathbb{R})}\|\partial_{ttx}u\|_{L^\infty(\Omega_T)}$$
$$+\|f'''\|_{L^\infty(\mathbb{R})}\|\partial_t u\|^2_{L^\infty(\Omega_T)}\|\partial_x u\|_{L^\infty(\Omega_T)}+\frac{1}{2\varepsilon^2}\|\partial_{ttx}u\|_{L^\infty(\Omega_T)}$$
$$+\|f''\|_{L^\infty(\mathbb{R})}\|\partial_t u\|_{L^\infty(\Omega_T)}\|\partial_{tx}u\|_{L^\infty(\Omega_T)}\Big]$$

$$\lesssim h^2+\frac{\tau^2}{\varepsilon^6}+\frac{\tau^2}{\varepsilon^4}\lesssim h^2+\frac{\tau^2}{\varepsilon^6},\quad 1\le n\le\frac{T}{\tau}. \tag{I.10}$$

These immediately imply the estimates in (I.5). □

Lemma 4 *There exists $h_0>0$ and $\tau_0>0$ sufficiently small, under the assumption* (*B2*) *and when $0<\tau\le\tau_0$ and $0<h\le h_0$, there exists a unique solution $u_j^n(j=0,1,\ldots,M;\, n\ge 0)$ of the problem* (2.2) *with* (2.7) *and* (2.8).

Proof The argument follows the analogous lines as in [15,48] for the standard KG equation, i.e. $\varepsilon=1$ in (2.1), and we omit the details here for brevity. □

Lemma 5 *For $j=0,1,\ldots,M,\ n\ge 1$, denote*

$$\eta_j^n=G(u(x_j,t_{n+1}),u(x_j,t_{n-1}))-G\left(u_j^{n+1},u_j^{n-1}\right), \tag{I.11}$$

assume $\tau\lesssim h$ and under the assumptions (*A*) *and* (*B2*), *we have*

$$\|\eta^n\|^2_{l^2}\lesssim\|e^{n-1}\|^2_{l^2}+\|e^{n+1}\|^2_{l^2},\quad n\ge 1, \tag{I.12}$$
$$\|\delta_x^+\eta^n\|^2_{l^2}\lesssim\|e^{n-1}\|^2_{l^2}+\|\delta_x^+e^{n-1}\|^2_{l^2}+\|e^{n+1}\|^2_{l^2}+\|\delta_x^+e^{n+1}\|^2_{l^2}. \tag{I.13}$$

Proof From (I.11), noticing (2.6) and the assumption (B2), we get

$$|\eta_j^n|=\left|\int_0^1\left[f\left(\theta u(x_j,t_{n+1})+(1-\theta)u(x_j,t_{n-1})\right)-f\left(\theta u_j^{n+1}+(1-\theta)u_j^{n-1}\right)\right]d\theta\right|$$
$$\le\|f'\|_{L^\infty(\mathbb{R})}\int_0^1\left[\theta\left|u(x_j,t_{n+1})-u_j^{n+1}\right|+(1-\theta)\left|u(x_j,t_{n-1})-u_j^{n-1}\right|\right]d\theta$$
$$\lesssim\left|e_j^{n-1}\right|+\left|e_j^{n+1}\right|,\ j=0,1,\ldots,M;\quad n\ge 1.$$

Using Hölder inequality, we get (I.12) immediately. Similarly, for $j = 0, 1, \ldots, M-1$ and $n \geq 1$, we can obtain

$$|\delta_x^+ \eta_j^n| \lesssim |e_j^{n-1}| + |\delta_x^+ e_j^{n-1}| + |e_{j+1}^{n-1}| + |e_j^{n+1}| + |\delta_x^+ e_j^{n+1}| + |e_{j+1}^{n+1}|.$$

This together with the Hölder inequality implies (I.13) immediately. □

Combining Lemmas 3 and 5, we give the proof of Theorem 2:

Proof of Theorem 2 Subtracting (2.2) and (2.8) from (I.4), noticing (2.7) and (2.24), we see the error e_j^n satisfies

$$\varepsilon^2 \delta_t^2 e_j^n - \frac{1}{2}\left(\delta_x^2 e_j^{n+1} + \delta_x^2 e_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(e_j^{n+1} + e_j^{n-1}\right) = \xi_j^n - \eta_j^n, \tag{I.14a}$$

$$e_0^n = e_M^n, \quad e_{-1}^n = e_{M-1}^n, \quad n = 0, 1, \ldots, \tag{I.14b}$$

$$e_j^0 = 0, \quad e_j^1 = \tau \xi_j^0, \quad j = 0, 1, \ldots, M. \tag{I.14c}$$

Define the 'energy' for the error vector $e^n (n = 0, 1, \ldots)$ as

$$\mathcal{E}^n = \varepsilon^2 \left\|\delta_t^+ e^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^+ e^n\right\|_{l^2}^2 + \left\|\delta_x^+ e^{n+1}\right\|_{l^2}^2\right) + \frac{1}{2\varepsilon^2}\left(\left\|e^n\right\|_{l^2}^2 + \left\|e^{n+1}\right\|_{l^2}^2\right). \tag{I.15}$$

Multiplying both sides of (I.14a) by $h\left(e_j^{n+1} - e_j^{n-1}\right)$, then summing up for $j = 0, 1, \ldots, M-1$, noticing (I.1) and (I.15), we get

$$\mathcal{E}^n - \mathcal{E}^{n-1} = h \sum_{j=0}^{M-1} \left(\xi_j^n - \eta_j^n\right)\left(e_j^{n+1} - e_j^{n-1}\right), \quad n \geq 1. \tag{I.16}$$

From (I.16), using Young's inequality, noticing Lemma 5, we have

$$\begin{aligned}
\mathcal{E}^n - \mathcal{E}^{n-1} &\leq h \sum_{j=0}^{M-1} \left(\left|\xi_j^n\right| + \left|\eta_j^n\right|\right)\left|e_j^{n+1} - e_j^{n-1}\right| \\
&= \tau h \sum_{j=0}^{M-1} \left(\left|\xi_j^n\right| + \left|\eta_j^n\right|\right)\left|\delta_t^+ e_j^n + \delta_t^+ e_j^{n-1}\right| \\
&\leq \tau \left[\frac{1}{\varepsilon^2}\left(\|\xi^n\|_{l^2}^2 + \|\eta^n\|_{l^2}^2\right) + \varepsilon^2\left(\|\delta_t^+ e^n\|_{l^2}^2 + \|\delta_t^+ e^{n-1}\|_{l^2}^2\right)\right] \\
&\lesssim \tau\left(\mathcal{E}^n + \mathcal{E}^{n-1}\right) + \frac{\tau}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad n \geq 1.
\end{aligned} \tag{I.17}$$

Thus, there exists a constant $\tau_0 > 0$ sufficiently small and independent of ε and h, such that when $0 < \tau \le \tau_0$

$$\mathcal{E}^n - \mathcal{E}^{n-1} \lesssim \tau \mathcal{E}^{n-1} + \frac{\tau}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad n \ge 1. \tag{I.18}$$

Summing the above inequality up for n, we get

$$\mathcal{E}^n - \mathcal{E}^0 \lesssim \tau \sum_{m=0}^{n-1} \mathcal{E}^m + \frac{T}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 1 \le n \le \frac{T}{\tau} - 1. \tag{I.19}$$

Using the discrete Gronwall's inequality [30,38], we obtain

$$\mathcal{E}^n \lesssim \mathcal{E}^0 + \frac{T}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 1 \le n \le \frac{T}{\tau} - 1. \tag{I.20}$$

Combining (I.14), (I.15) for $n = 0$ and (I.5), we have

$$\begin{aligned}\mathcal{E}^0 &= \varepsilon^2 \|\xi^0\|_{l^2}^2 + \frac{\tau^2}{2}\|\delta_x^+ \xi^0\|_{l^2}^2 + \frac{\tau^2}{\varepsilon^2}\|\xi^0\|_{l^2}^2 \\ &\lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2 \left(\varepsilon^2 + \frac{\tau^2}{2} + \frac{\tau^2}{\varepsilon^2}\right) \lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2 \left(1 + \frac{\tau^2}{\varepsilon^2}\right).\end{aligned} \tag{I.21}$$

Plugging (I.21) into (I.20), we get

$$\mathcal{E}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \le n \le \frac{T}{\tau} - 1. \tag{I.22}$$

In addition, define another 'energy' for the error vector $e^n (n = 0, 1, \ldots)$ as

$$\begin{aligned}\hat{\mathcal{E}}^n &= \varepsilon^2 \left\|\delta_x^+ \delta_t^+ e^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^2 e^n\right\|_{l^2}^2 + \left\|\delta_x^2 e^{n+1}\right\|_{l^2}^2\right) \\ &\quad + \frac{1}{2\varepsilon^2}\left(\left\|\delta_x^+ e^n\right\|_{l^2}^2 + \left\|\delta_x^+ e^{n+1}\right\|_{l^2}^2\right).\end{aligned} \tag{I.23}$$

Multiplying both sides of (I.14a) by $h\left(\delta_x^2 e_j^{n+1} - \delta_x^2 e_j^{n-1}\right)$, similar to the above procedure, we can obtain

$$\hat{\mathcal{E}}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \le n \le \frac{T}{\tau} - 1. \tag{I.24}$$

Combining (I.15), (I.22), (I.23), and (I.24), noticing that $\|e^n\|_{l^2}^2 + \|e^{n+1}\|_{l^2}^2 \le 2\varepsilon^2 \mathcal{E}^n$ and $\|\delta_x^+ e^n\|_{l^2}^2 + \|\delta_x^+ e^{n+1}\|_{l^2}^2 \le 2\varepsilon^2 \hat{\mathcal{E}}^n$ when $0 < \varepsilon \le 1$, we immediately obtain the error estimate in (2.25). □

Appendix II: Proof of Theorem 3 for Expt-FD (2.5)

Proof Define

$$\tilde{\xi}_j^n := \varepsilon^2 \delta_t^2 (u(x_j, t_n)) - \delta_x^2 (u(x_j, t_n)) + \frac{1}{\varepsilon^2} u(x_j, t_n) + f(u(x_j, t_n)), \quad \text{(II.1)}$$

$$\tilde{\eta}_j^n := f(u(x_j, t_n)) - f(u_j^n), \quad j = 0, 1, \ldots, M-1, \quad n \geq 1. \quad \text{(II.2)}$$

Similar to Lemmas 3 and 5, we can prove

$$\left\| \tilde{\xi}^n \right\|_{l^2} + \left\| \delta_x^+ \tilde{\xi}^n \right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \leq n \leq \frac{T}{\tau}; \quad \|\delta_x^2 \tilde{\xi}^0\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad \text{(II.3)}$$

$$\|\tilde{\eta}^n\|_{l^2}^2 \lesssim \|e^n\|_{l^2}^2, \quad \|\delta_x^+ \tilde{\eta}^n\|_{l^2}^2 \lesssim \|e^n\|_{l^2}^2 + \|\delta_x^+ e^n\|_{l^2}^2, \quad n \geq 1. \quad \text{(II.4)}$$

Subtracting (2.5) from (II.1), noticing (2.7), (2.8) and (II.2), we get

$$\varepsilon^2 \delta_t^2 e_j^n - \delta_x^2 e_j^n + \frac{1}{\varepsilon^2} e_j^n = \tilde{\xi}_j^n - \tilde{\eta}_j^n, \quad \text{(II.5a)}$$

$$e_0^n = e_M^n, \quad e_{-1}^n = e_{M-1}^n, \quad n = 0, 1, \ldots, \quad \text{(II.5b)}$$

$$e_j^0 = 0, \quad e_j^1 = \tau \xi_j^0, \quad j = 0, 1, \ldots, M. \quad \text{(II.5c)}$$

Define the ‘energy’ for the error vector $e^n (n = 0, 1, \ldots)$ as

$$\mathcal{S}^n := \left(\varepsilon^2 - \frac{\tau^2}{2\varepsilon^2} - \frac{\tau^2}{h^2} \right) \|\delta_t^+ e^n\|_{l^2}^2 + \frac{1}{2\varepsilon^2} \left(\|e^{n+1}\|_{l^2}^2 + \|e^n\|_{l^2}^2 \right)$$
$$+ \frac{1}{2h} \sum_{j=0}^{M-1} \left[\left(e_{j+1}^{n+1} - e_j^n \right)^2 + \left(e_{j+1}^n - e_j^{n+1} \right)^2 \right], \quad n \geq 0. \quad \text{(II.6)}$$

Similar to the proof in Theorem 2, with the help of (I.2) and (I.3), noticing (II.5), (2.5), (2.7), (2.8) and (II.6), in view of the estimates (II.3), we obtain

$$\mathcal{S}^n \lesssim \mathcal{S}^0 + \frac{1}{\varepsilon^2} \left(h^2 + \frac{\tau^2}{\varepsilon^6} \right)^2, \quad 0 \leq n \leq \frac{T}{\tau} - 1. \quad \text{(II.7)}$$

Plugging (II.5) into (II.6) with $n = 0$, we get

$$\mathcal{S}^0 \lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6} \right)^2 \left(1 + \frac{\tau^2}{\varepsilon^2} \right). \quad \text{(II.8)}$$

Similarly, define another 'energy' as

$$\hat{\mathcal{S}}^n := \left(\varepsilon^2 - \frac{\tau^2}{2\varepsilon^2} - \frac{\tau^2}{h^2}\right) \|\delta_x^+\delta_t^+ e^n\|_{l^2}^2 + \frac{1}{2\varepsilon^2}\left(\left\|\delta_x^+ e^{n+1}\right\|_{l^2}^2 + \|\delta_x^+ e^n\|_{l^2}^2\right)$$
$$+\frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ e_{j+1}^{n+1} - \delta_x^+ e_j^n\right)^2 + \left(\delta_x^+ e_{j+1}^n - \delta_x^+ e_j^{n+1}\right)^2\right], \quad n \geq 0, \tag{II.9}$$

we can obtain

$$\mathcal{S}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \leq n \leq \frac{T}{\tau} - 1. \tag{II.10}$$

Thus (2.27) is a combination of (II.6)–(II.10) by noticing $\|e^n\|_{l^2}^2 + \|e^{n+1}\|_{l^2}^2 \leq 2\varepsilon^2\mathcal{S}^n$, $\|\delta_x^+ e^n\|_{l^2}^2 + \|\delta_x^+ e^{n+1}\|_{l^2}^2 \leq 2\varepsilon^2\hat{\mathcal{S}}^n$ and $0 < \varepsilon \leq 1$. □

Appendix III: Proof of Theorem 7 for Gautschi-FP in linear case

Proof From (3.23), we have $u_I(x_j, 0) = (I_M\phi)(x_j) = \phi(x_j) = u_j^0$ for $j = 0, 1, \ldots, M$. Thus (3.25) is valid for $n = 0$. From (3.5) and (3.21), when $\alpha^n = \alpha$ for $n \geq 0$, we get

$$\beta_l^n = \beta_l, \quad n \geq 0, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1. \tag{III.1}$$

Plugging (3.13) into (3.15) with $n = 0$, noticing (III.1) and (3.24), we get

$$\begin{aligned} u_j^1 &= \sum_{l=-M/2}^{M/2-1}\left[p_l^0\widetilde{\phi}_l + q_l^0\widetilde{\gamma}_l + r_l^0\alpha\widetilde{\phi}_l\right] e^{2ijl\pi/M} \\ &= \sum_{l=-M/2}^{M/2-1}\left[\widetilde{\phi}_l\cos(t_1\beta_l) + \widetilde{\gamma}_l\frac{\sin(t_1\beta_l)}{\varepsilon^2\beta_l}\right] e^{2ijl\pi/M} \\ &= u_I(x_j, t_1), \quad j = 0, 1, \ldots, M. \end{aligned} \tag{III.2}$$

Thus (3.25) is valid for $n = 1$. Assume (3.25) is valid for $n = 0, 1, \ldots, m$. When $n = m + 1$, from (3.15) with $n = m$, noticing (3.14) and (III.1), we have

$$\widetilde{(u^{m+1})}_l = -\widetilde{(u^{m-1})}_l + p_l^m\,\widetilde{(u^m)}_l + r_l^m\,\alpha\widetilde{(u^m)}_l = -\widetilde{(u^{m-1})}_l + 2\cos(\tau\beta_l)\,\widetilde{(u^m)}_l$$
$$= -\left[\widetilde{\phi}_l\cos(t_{m-1}\beta_l) + \widetilde{\gamma}_l\frac{\sin(t_{m-1}\beta_l)}{\varepsilon^2\beta_l}\right] + 2\cos(\tau\beta_l)$$

$$\times \left[\widetilde{\phi}_l \cos(t_m \beta_l) + \widetilde{\gamma}_l \frac{\sin(t_m \beta_l)}{\varepsilon^2 \beta_l} \right]$$
$$= \widetilde{\phi}_l \cos(t_{m+1} \beta_l) + \widetilde{\gamma}_l \frac{\sin(t_{m+1} \beta_l)}{\varepsilon^2 \beta_l}, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1.$$

Plugging the above equality into (3.15) with $n = m$ and noticing (3.24) with $t = t_{m+1}$, we obtain (3.25) for $n = m + 1$, thus the claim (3.25) is verified by mathematical induction. From (3.25), noticing (3.20) and (3.24), we obtain

$$\|e^n(x)\|_{L^2}^2 \lesssim \|\phi - I_M \phi\|_{L^2}^2 + \|\gamma - I_M \gamma\|_{L^2}^2 \lesssim h^{2m},$$
$$\|\nabla e^n(x)\|_{L^2}^2 \lesssim \|\nabla(\phi - I_M \phi)\|_{L^2}^2 + \|\nabla(\gamma - I_M \gamma)\|_{L^2}^2 \lesssim h^{2(m-1)},$$

which complete the proof of (3.26). □

Appendix IV: Proof of Theorem 8 for Gautschi-FD in linear case

Proof Let $u_j(t)$ be the solution of (3.17) with the initial condition

$$u_j(0) = \phi(x_j), \quad \frac{\mathrm{d}}{\mathrm{d}t} u_j(0) = \frac{1}{\varepsilon^2} \gamma(x_j), \quad j = 0, 1, \ldots, M.$$

Similar to the proof of Theorem 7, we have for $0 \le j \le M, \ n \ge 0$

$$u_j^n = u_j(t_n) = \sum_{l=-M/2}^{M/2-1} \left[\widetilde{\phi}_l \cos(t \beta_l^h) + \widetilde{\gamma}_l \frac{\sin(t \beta_l^h)}{\varepsilon^2 \beta_l^h} \right] \mathrm{e}^{i \mu_l (x_j - a)},$$

where

$$\beta_l^h = \frac{1}{\varepsilon^2} \sqrt{1 + \varepsilon^2 \left(\lambda_l^2 + \alpha \right)} \ge \frac{1}{\sqrt{2} \varepsilon^2}, \quad l = 0, \pm 1, \ldots. \tag{IV.1}$$

Let

$$e_j(t) = u(x_j, t) - u_j(t), \quad j = 0, 1, \ldots, M,$$
$$\xi_j(t) = \varepsilon^2 \tfrac{\mathrm{d}^2}{\mathrm{d}t^2} u(x_j, t) - \delta_x^2 u(x_j, t) + \left(\tfrac{1}{\varepsilon^2} + \alpha \right) u(x_j, t) = \tfrac{h^2}{12} \partial_{xxxx} u(\tilde{x}_j(t), t), \tag{IV.2}$$

where $\tilde{x}_j(t)$ is located between x_{j-1} and x_{j+1}. Subtracting (3.17) from (IV.2), we have

$$\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} e_j(t) - \delta_x^2 e_j(t) + \left(\frac{1}{\varepsilon^2} + \alpha \right) e_j(t) = \xi_j(t), \quad j = 0, 1, \ldots, M, \ t \ge 0, \tag{IV.3}$$
$$e_0(t) = e_M(t), \quad e_{-1}(t) = e_{M-1}(t), \quad e_j(0) = 0, \quad \frac{\mathrm{d}}{\mathrm{d}t} e_j(0) = 0.$$

Taking the discrete Fourier transform of (IV.3), we get

$$\begin{aligned}&\varepsilon^2\frac{\mathrm{d}^2}{\mathrm{d}t^2}\widetilde{e}_l(t)+(\varepsilon\beta_l^h)^2\widetilde{e}_l(t)=\widetilde{\xi}_l(t),\\&\widetilde{e}_l(0)=0,\quad \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{e}_l(0)=0,\quad l=-\frac{M}{2},\ldots,\frac{M}{2}-1.\end{aligned}\tag{IV.4}$$

Solving the above ODEs, we have

$$\widetilde{e}_l(t)=\frac{1}{\beta_l^h\varepsilon^2}\int_0^t\sin(\beta_l^h(t-s))\widetilde{\xi}_l(s)\mathrm{d}s,\quad l=-\frac{M}{2},\ldots,\frac{M}{2}-1.\tag{IV.5}$$

Plugging (IV.2) into (IV.5), noticing $\phi,\ \gamma\in W_p^{4,\infty}(\Omega)$ and (3.20), using the Hölder's inequality and Parseval's identity, we obtain

$$\begin{aligned}\sum_{l=-M/2}^{M/2-1}|\widetilde{e}_l(t)|^2&\leq 2\sum_{l=-M/2}^{M/2-1}\left[\int_0^t\left|\widetilde{\xi}_l(s)\right|\mathrm{d}s\right]^2\leq 2t\int_0^t\sum_{l=-M/2}^{M/2-1}\left|\widetilde{\xi}_l(s)\right|^2\mathrm{d}s\\&\leq\frac{2t}{M}\int_0^t\sum_{j=0}^{M-1}|\xi_j(t)|^2\mathrm{d}s\leq\frac{2T}{M}\int_0^T\sum_{j=0}^{M-1}|\xi_j(s)|^2\mathrm{d}s\lesssim h^4,\quad 0\leq t\leq T.\end{aligned}$$

Noticing $e_j^n=e_j(t_n)(j=0,1,\ldots,M,0\leq n\leq T/\tau)$ and using the Parseval's equality, we obtain the estimate (3.27) immediately. □

Appendix V: Proof of Theorem 9 for Gautschi-FP in nonlinear case

Proof We will prove (3.32) by the method of mathematical induction [4]. From the discretization of the initial data, i.e. $u_M^0=P_M\phi$, we have

$$\begin{aligned}&\|u(x,t=0)-u_M^0\|_{L^2}=\|\phi-P_M\phi\|_{L^2}\lesssim h^{m_0},\\&\|\nabla[u(x,t=0)-u_M^0]\|_{L^2}=\|\nabla\phi-P_M\nabla\phi\|_{L^2}\lesssim h^{m_0-1},\\&\|u_M^0\|_{L^\infty}\leq\|P_M\phi-\phi\|_{L^\infty}+\|\phi\|_{L^\infty}\leq Ch^{m_0-1}+M_1.\end{aligned}$$

Thus there exists a $h_1>0$ sufficiently small and independent of ε such that, when $0<h\leq h_1$, the three inequalities in (3.32) are valid for $n=0$.

Denote the 'error' function

$$e^n(x):=P_Mu(x,t_n)-u_M^n(x)=\sum_{l=-M/2}^{M/2-1}\widehat{e}_l^{\,n}e^{i\mu_l(x-a)},\quad a\leq x\leq b,\tag{V.1}$$

then we have

$$\widehat{e}_l^n = \widehat{u}_l(t_n) - \widehat{(u_M^n)}_l, \quad l = -\frac{M}{2}, \ldots, \frac{M}{2} - 1, \quad n \geq 0, \tag{V.2}$$

with $\widehat{u}_l(t_n)(l = 0, \pm 1, \ldots)$ the Fourier coefficients of $u(x, t_n)$. Using the triangle inequality and Parseval's equality, we get

$$\begin{aligned} \|u(x, t_n) - u_M^n(x)\|_{L^2} &\leq \|u(x, t_n) - P_M u(x, t_n)\|_{L^2} + \|e^n(x)\|_{L^2} \\ &\lesssim h^{m_0} + \sqrt{\sum_{l=-M/2}^{M/2-1} |\widehat{e}_l^n|^2}, \quad 0 \leq n \leq \frac{T}{\tau}. \end{aligned} \tag{V.3}$$

Similarly, we have

$$\|\nabla[u(x, t_n) - u_M^n(x)]\|_{L^2} \lesssim h^{m_0 - 1} + \sqrt{\sum\nolimits_{l=-M/2}^{M/2-1} \mu_l^2 |\widehat{e}_l^n|^2}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{V.4}$$

Thus we only need to estimate the last terms in the above two inequalities.

Similar to the derivation in (3.3)–(3.9), for $l = 0, \pm 1, \ldots$, we have

$$\widehat{u}_l(\tau) = \widehat{\phi}_l \cos(\tau \beta_l^0) + \widehat{\gamma}_l \frac{\sin(\tau \beta_l^0)}{\varepsilon^2 \beta_l^0} - \frac{1}{\varepsilon^2 \beta_l^0} \int_0^\tau \widehat{G}_l^0(s) \sin(\beta_l^0(\tau - s)) \mathrm{d}s, \tag{V.5}$$

$$\begin{aligned} \widehat{u}_l(t_{n+1}) &= -\widehat{u}_l(t_{n-1}) + 2\cos(\tau \beta_l^n) \widehat{u}_l(t_n) \\ &\quad - \frac{1}{\varepsilon^2 \beta_l^n} \int_0^\tau \left[\widehat{G}_l^n(-s) + \widehat{G}_l^n(s)\right] \sin(\beta_l^n(\tau - s)) \mathrm{d}s, \quad n \geq 1, \end{aligned} \tag{V.6}$$

where

$$\widehat{G}_l^n(s) = \widehat{(f(u))}_l(t_n + s) - \alpha^n \widehat{u}_l(t_n + s), \quad s \in \mathbb{R}, \quad n \geq 0. \tag{V.7}$$

For each $l = -M/2, \ldots, M/2 - 1$, subtracting (3.12) and (3.11) from (V.6) and (V.5), respectively, we obtain the equation for the 'error' function $\widehat{e}_l^n$ as

$$\widehat{e}_l^{n+1} = -\widehat{e}_l^{n-1} + 2\cos(\beta_l^n \tau) \widehat{e}_l^n + \widehat{\xi}_l^n, \quad 1 \leq n \leq \frac{T}{\tau} - 1, \tag{V.8}$$

$$\widehat{e}_l^0 = 0, \quad \widehat{e}_l^1 = \widehat{\xi}_l^0, \tag{V.9}$$

where

$$\widehat{\xi}_l^n = \frac{1}{\varepsilon^2 \beta_l^n} \int_0^\tau \widehat{W}_l^n(s) \sin(\beta_l^n(\tau - s)) ds, \quad 0 \leq n \leq \frac{T}{\tau} - 1, \tag{V.10}$$

with for $0 \leq s \leq \tau$

$$\widehat{W}_l^n(s) = \begin{cases} \widehat{f(\phi)}_l - \alpha^0 \widehat{\phi}_l - \widehat{G}_l^0(s), & n = 0, \\ 2\widehat{f(u_M^n)}_l - 2\alpha^n \widehat{(u_M^n)}_l - \widehat{G}_l^n(-s) - \widehat{G}_l^n(s), & 1 \leq n \leq \frac{T}{\tau} - 1. \end{cases} \tag{V.11}$$

Combining (3.16), (3.29) and (3.32) with $n = 0$, noticing (3.5), under the condition (3.31), we get

$$\begin{aligned} &0 \leq \alpha^0 \leq M_3, \quad \varepsilon^2 \beta_l^0 \geq 1, \quad 0 < \tau \beta_l^0 \leq \frac{\pi}{3}, \quad \frac{1}{2} \leq \cos(\beta_l^0 \tau) < 1, \\ &0 \leq \sin(\beta_l^0(\tau - s)) \leq \sin(\beta_l^0 \tau) < 1, \quad 0 \leq s \leq \tau. \end{aligned}$$

From (V.10) with $n = 0$, using the Hölder inequality, we obtain

$$\begin{aligned} \left|\widehat{\xi}_l^0\right|^2 &= \left| \frac{1}{\varepsilon^2 \beta_l^0} \int_0^\tau \widehat{W}_l^0(s) \sin(\beta_l^0(\tau - s)) \mathrm{d}s \right|^2 \\ &\leq \int_0^\tau \sin(\beta_l^0(\tau - s)) \mathrm{d}s \cdot \int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \sin(\beta_l^0(\tau - s)) \mathrm{d}s \\ &\leq \tau \left[1 - \cos(\beta_l^0 \tau)\right] \frac{\sin(\beta_l^0 \tau)}{\beta_l^0 \tau} \int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s \\ &\leq \tau \left[1 - \cos(\beta_l^0 \tau)\right] \int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s. \end{aligned} \tag{V.12}$$

Summing the above inequality for $l = -M/2, \ldots, M/2-1$, noticing (V.9) and (V.12), we obtain

$$\begin{aligned} \|e^1\|_{L^2}^2 &= (b - a) \sum_{l=-M/2}^{M/2-1} \left|\widehat{e}_l^1\right|^2 = (b - a) \sum_{l=-M/2}^{M/2-1} \left|\widehat{\tau}_l^0\right|^2 \\ &\leq \tau (b - a) \sum_{l=-M/2}^{M/2-1} \int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s. \end{aligned}$$

Plugging (V.11), (V.7) and (3.5) into the above inequality, using the triangle inequality and Parseval's equality, we get

$$
\begin{aligned}
\|e^1\|_{L^2}^2 &\le \tau(b-a) \sum_{l=-M/2}^{M/2-1} \int_0^\tau \left| \widehat{(f(\phi))}_l - \widehat{(f(u))}_l(s) + \alpha^0(\widehat{u}_l(s) - \widehat{\phi}_l) \right|^2 \mathrm{d}s \\
&\le 2\tau(b-a) \int_0^\tau \sum_{l=-M/2}^{M/2-1} \left[\left| \widehat{(f(\phi))}_l - \widehat{(f(u))}_l(s) \right|^2 + (\alpha^0)^2 \left| \widehat{u}_l(s) - \widehat{\phi}_l \right|^2 \right] \mathrm{d}s \\
&= 2\tau \int_0^\tau \left(\| P_M[f(u(\cdot,s)) - f(\phi)] \|_{L^2}^2 + (\alpha^0)^2 \| P_M[u(\cdot,s) - \phi] \|_{L^2}^2 \right) \mathrm{d}s \\
&\le 2\tau \int_0^\tau \left(\| f(u(\cdot,s)) - f(\phi) \|_{L^2}^2 + M_3^2 \| u(\cdot,s) - \phi \|_{L^2}^2 \right) \mathrm{d}s. \qquad \text{(V.13)}
\end{aligned}
$$

Under the assumption on u, using the Hölder inequality, we get

$$
\begin{aligned}
\|u(\cdot,s) - \phi\|_{L^2}^2 &= \int_a^b |u(x,s) - u(x,0)|^2 \, dx = \int_a^b \left| \int_0^s \partial_w u(x,w) \, dw \right|^2 dx \\
&\le \int_a^b s \int_0^s |\partial_w u(x,w)|^2 \, dw \, dx = s \int_0^s \|\partial_w u(\cdot,w)\|_{L^2}^2 \, dw \\
&\le s^2 \|\partial_t u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2 \lesssim \frac{s^2}{\varepsilon^4}, \quad 0 \le s \le \tau. \qquad \text{(V.14)}
\end{aligned}
$$

Similarly, under the assumption on u and f, we have

$$
\|f(u(\cdot,s)) - f(\phi)\|_{L^2}^2 \le s^2 M_2^2 \|\partial_t u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2 \lesssim \frac{s^2}{\varepsilon^4}, \quad 0 \le s \le \tau. \qquad \text{(V.15)}
$$

Plugging (V.14) and (V.15) into (V.13), noticing (V.3) with $n = 1$, we obtain

$$
\|e^1\|_{L^2}^2 \lesssim \tau \int_0^\tau \frac{s^2}{\varepsilon^4} \, ds \lesssim \frac{\tau^4}{\varepsilon^4} \lesssim \frac{\tau^4}{\varepsilon^8} \;\Rightarrow\; \|u(x,t_1) - u_M^1(x)\|_{L^2} \lesssim h^{m_0} + \frac{\tau^2}{\varepsilon^4}.
$$

Similarly, we can get

$$
\|\nabla[u(x,t_1) - u_M^1(x)]\|_{L^2} \lesssim h^{m_0 - 1} + \frac{\tau^2}{\varepsilon^4}.
$$

This, together with the triangle inequality and inverse inequality, implies

$$\begin{aligned}
\|u_M^1\|_{L^\infty} - M_1 &\le \|u_M^1\|_{L^\infty} - \|u(x,t_1)\|_{L^\infty} \le \|u_M^1 - u(x,t_1)\|_{L^\infty} \\
&\le \|P_M u(x,t_1) - u(x,t_1)\|_{L^\infty} + \|u_M^1(x) - P_M u(x,t_1)\|_{L^\infty} \\
&\lesssim \|u(x,t_1) - P_M u(x,t_1)\|_{L^\infty} + \frac{\|u_M^1(x) - P_M u(x,t_1)\|_{H^1}}{C_d(h)} \\
&\lesssim h^{m_0-1} + \frac{1}{C_d(h)}\|e^1\|_{H^1} \\
&\lesssim h^{m_0-1}\left(1 + \frac{1}{C_d(h)}\right) + \frac{\tau^2}{\varepsilon^4 C_d(h)}. \qquad \text{(V.16)}
\end{aligned}$$

Thus under the assumption $\tau \lesssim \varepsilon^2\sqrt{C_d(h)}$, there exist $h_2 > 0$ and $\tau_2 > 0$ sufficiently small and independent of ε, such that when $0 < h \le h_2$ and $0 < \tau \le \tau_2$, we have

$$\|u_M^1\|_{L^\infty} \le 1 + M_1.$$

Therefore, the three inequalities in (3.32) are valid when $n = 1$.

Now we assume that (3.32) is valid for all $1 \le n \le m - 1 \le \frac{T}{\tau} - 1$, then we need to show that it is still valid when $n = m$. Denote

$$\mathcal{E}^n = \sum_{l=-M/2}^{M/2-1} \widehat{\mathcal{E}}_l^n, \quad \widehat{\mathcal{E}}_l^n = \left|\widehat{e}_l^{n+1}\right|^2 + \left|\widehat{e}_l^n\right|^2 + \frac{\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)}\left|\widehat{e}_l^{n+1} - \widehat{e}_l^n\right|^2. \qquad \text{(V.17)}$$

For each $l = -M/2, \ldots, M/2-1$ and $1 \le n \le m-1$, noticing (3.5), under the condition (3.31), we get

$$\begin{aligned}
&0 \le \alpha^{n-1} \le \alpha^n \le M_3, \quad 1 \le \varepsilon^2\beta_l^{n-1} \le \varepsilon^2\beta_l^n, \quad 0 < \tau\beta_l^{n-1} \le \tau\beta_l^n \le \frac{\pi}{3}, \\
&\frac{1}{2} \le \cos(\beta_l^n\tau) \le \cos(\beta_l^{n-1}\tau) < 1, \quad \frac{\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)} \le \frac{\cos(\beta_l^{n-1}\tau)}{1-\cos(\beta_l^{n-1}\tau)}, \\
&0 \le \sin(\beta_l^n(\tau - s)) \le \sin(\beta_l^n\tau) < 1, \quad 0 \le s \le \tau.
\end{aligned}$$

Then similar to (V.12), we obtain

$$\left|\widehat{\xi}_l^n\right|^2 \le \tau\left[1 - \cos(\beta_l^n\tau)\right]\int_0^\tau \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s.$$

Multiplying both sides of (V.8) by $\overline{\widehat{e}_l^{n+1}} - \overline{\widehat{e}_l^{n-1}}$ (here $\overline{w}$ denotes the conjugate of w) and adding with its conjugate, then dividing by $1-\cos(\beta_l^n\tau)$, we have

$$\begin{aligned}
\widehat{\mathcal{E}}_l^n - \widehat{\mathcal{E}}_l^{n-1} &\le \frac{1}{1-\cos(\beta_l^n\tau)} \left|\widehat{\xi}_l^n\right| \cdot \left|\widehat{e}_l^{n+1} - \widehat{e}_l^{n-1}\right| \\
&\le \frac{1}{1-\cos(\beta_l^n\tau)} \left(2\tau \left|\widehat{e}_l^{n+1} - \widehat{e}_l^{n}\right|^2 + 2\tau \left|\widehat{e}_l^{n} - \widehat{e}_l^{n-1}\right|^2 + \frac{1}{\tau}\left|\widehat{\xi}_l^n\right|^2\right) \\
&\le \frac{4\tau\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)} \left(\left|\widehat{e}_l^{n+1} - \widehat{e}_l^{n}\right|^2 + \left|\widehat{e}_l^{n} - \widehat{e}_l^{n-1}\right|^2\right) + \int_0^\tau \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s \\
&\le 4\tau\left(\widehat{\mathcal{E}}_l^n + \widehat{\mathcal{E}}_l^{n-1}\right) + \int_0^\tau \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s.
\end{aligned}$$

Summing the above inequality for $l = -M/2, \ldots, M/2-1$, we obtain

$$\mathcal{E}^n - \mathcal{E}^{n-1} \le 4\tau\left(\mathcal{E}^n + \mathcal{E}^{n-1}\right) + \int_0^\tau \sum_{l=-M/2}^{M/2-1} \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s, \quad 0 \le n \le m-1.$$

Summing the above inequality for $n = 1, 2, \ldots, m-1$, we get, when $\tau \le 1/8$

$$\mathcal{E}^{m-1} \le 2\mathcal{E}^0 + 8\tau\sum_{n=0}^{m-2}\mathcal{E}^n + 2\sum_{n=1}^{m-2}\int_0^\tau \sum_{l=-M/2}^{M/2-1} \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s, \quad 2 \le m \le \frac{T}{\tau}.$$

Using the discrete Gronwall's inequality, we get

$$\mathcal{E}^{m-1} \le C\left[\mathcal{E}^0 + \sum_{n=1}^{m-1}\int_0^\tau \sum_{l=-M/2}^{M/2-1} \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s\right], \quad 2 \le m \le \frac{T}{\tau}, \qquad \text{(V.18)}$$

where the constant C is independent of h (or l), τ (or m), and ε. Combining (V.1), (V.17) and (V.18), we obtain

$$\begin{aligned}
\|e^m\|_{L^2}^2 &= (b-a)\sum_{l=-M/2}^{M/2-1}\left|\widehat{e}_l^m\right|^2 \le (b-a)\mathcal{E}^{m-1} \\
&\le C(b-a)\left[\mathcal{E}^0 + \sum_{n=1}^{m-1}\int_0^\tau \sum_{l=-M/2}^{M/2-1} \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s\right]. \qquad \text{(V.19)}
\end{aligned}$$

From (V.17) with $n = 0$, noticing (V.9) and (V.12)–(V.15), we have

$$\mathcal{E}^0 = \sum_{l=-M/2}^{M/2-1} \frac{1}{1-\cos(\beta_l^0 \tau)} |\widehat{e}_l^{\,1}|^2 = \sum_{l=-M/2}^{M/2-1} \frac{1}{1-\cos(\beta_l^0 \tau)} |\widehat{\xi}_l^{\,0}|^2$$
$$= \tau \sum_{l=-M/2}^{M/2-1} \int_0^{\tau} \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s \lesssim \frac{\tau^4}{\varepsilon^4} \lesssim \frac{\tau^4}{\varepsilon^8}. \tag{V.20}$$

From (V.11), (V.7) and (3.5), using the triangle inequality, we get

$$\sum_{l=-M/2}^{M/2-1} \left|\widehat{W}_l^n(s)\right|^2 = \sum_{l=-M/2}^{M/2-1} \Big| 2\widehat{f(u_M^n)}_l - \widehat{(f(u))}_l(t_n - s) - \widehat{(f(u))}_l(t_n + s)$$
$$+\alpha^n \left[\widehat{u}_l(t_n - s) + \widehat{u}_l(t_n + s) - 2\widehat{(u_M^n)}_l\right]\Big|^2 \mathrm{d}s$$
$$\le \frac{2}{b-a} \|2f(u_M^n) - f(u(\cdot, t_n - s)) - f(u(\cdot, t_n + s))\|_{L^2}^2$$
$$+\frac{2M_3^2}{b-a} \|u(\cdot, t_n - s) + u(\cdot, t_n + s) - 2u_M^n\|_{L^2}^2. \tag{V.21}$$

Under the assumption on u, using the triangle inequality and Hölder inequality, noticing (3.32), we get

$$\left\|u(\cdot, t_n - s) + u(\cdot, t_n + s) - 2u_M^n\right\|_{L^2}^2$$
$$\le \|u(\cdot, t_n - s) + u(\cdot, t_n + s) - 2u(\cdot, t_n)\|_{L^2}^2 + 4\left\|u(\cdot, t_n) - u_M^n\right\|_{L^2}^2$$
$$\le \int_a^b \left| \int_0^s \int_{-w}^{w} \partial_{qq} u(x, t_n + q)\, dq\, dw \right|^2 dx + 4\left\|u(\cdot, t_n) - u_M^n\right\|_{L^2}^2$$
$$\le \int_0^s s \int_a^b \left| \int_{-w}^{w} \partial_{qq} u(x, t_n + q)\, dq \right|^2 dx\, dw + 4\left\|u(\cdot, t_n) - u_M^n\right\|_{L^2}^2$$
$$\le \int_0^s 2sw \int_{-w}^{w} \left\|\partial_{qq} u(\cdot, t_n + q)\right\|_{L^2}^2 dq\, dw + 4\left\|u(\cdot, t_n) - u_M^n\right\|_{L^2}^2$$
$$\le \frac{4s^4}{3} \|\partial_{tt} u(\cdot, t)\|_{L^\infty([0,T];L^2)}^2 + 4\left\|u(\cdot, t_n) - u_M^n\right\|_{L^2}^2$$
$$\lesssim \frac{\tau^4}{\varepsilon^8} + \left(\frac{\tau^2}{\varepsilon^4} + h^{m_0}\right)^2, \quad 0 \le s \le \tau, \quad 1 \le n \le m-1. \tag{V.22}$$

Similarly, under the assumption on u and f, we have

$$\begin{aligned}
&\|f(u(\cdot,t_n-s))+f(u(\cdot,t_n+s))-2f(u_M^n)\|_{L^2}^2 \\
&\le \frac{8s^4M_2^2}{3}\left[\|\partial_t u(\cdot,t)\|^4_{L^\infty([0,T];L^4)}+\|\partial_{tt}u(\cdot,t)\|^2_{L^\infty([0,T];L^2)}\right] \\
&\quad +4M_2^2\cdot\|u(\cdot,t_n)-u_M^n\|_{L^2}^2 \\
&\lesssim \frac{\tau^4}{\varepsilon^8}+\left(\frac{\tau^2}{\varepsilon^4}+h^{m_0}\right)^2, \quad 0\le s\le\tau, \quad 1\le n\le m-1. \qquad \text{(V.23)}
\end{aligned}$$

Plugging (V.23), (V.22), (V.21) and (V.20) into (V.19), we get

$$\begin{aligned}
\|e^m\|_{L^2}^2 &\lesssim \frac{\tau^4}{\varepsilon^8}+\tau(m-1)\left[\frac{\tau^4}{\varepsilon^8}+\left(\frac{\tau^2}{\varepsilon^4}+h^{m_0}\right)^2\right] \\
&\lesssim \frac{\tau^4}{\varepsilon^8}+T\left[\frac{\tau^4}{\varepsilon^8}+\left(\frac{\tau^2}{\varepsilon^4}+h^{m_0}\right)^2\right]\lesssim\left(\frac{\tau^2}{\varepsilon^4}+h^{m_0}\right)^2.
\end{aligned}$$

This, together with (V.3), implies that the first inequality in (3.32a) is valid for $n=m$. Similar to the above procedure by defining

$$\mathcal{S}^n=\sum_{l=-M/2}^{M/2-1}\mu_l^2\widehat{\mathcal{E}}_l^n, \quad n\ge 0,$$

and noticing

$$\begin{aligned}
\sum_{l=-M/2}^{M/2-1}\mu_l^2\left|\widehat{W}_l^n(s)\right|^2 &\lesssim \left\|\nabla\left[2f(u_M^n)-f(u(\cdot,t_n-s))-f(u(\cdot,t_n+s))\right]\right\|_{L^2}^2 \\
&\quad+\left\|\nabla\left[u(\cdot,t_n-s)+u(\cdot,t_n+s)-2u_M^n\right]\right\|_{L^2}^2,
\end{aligned}$$

we can obtain (3.32b). In addition, similar to the proof in (V.16), we have

$$\|u_M^m\|_{L^\infty}-M_1\lesssim h^{m_0-1}\left(1+\frac{1}{C_d(h)}\right)+\frac{\tau^2}{\varepsilon^4C_d(h)}.$$

Again under the assumption $\tau\lesssim\varepsilon^2\sqrt{C_d(h)}$, there exist $h_3>0$ and $\tau_3>0$ sufficiently small and independent of $2\le m\le T/\tau$, such that when $0<h\le h_3$ and $0<\tau\le\tau_3$, we have

$$\|u_M^m\|_{L^\infty}\le 1+M_1.$$

Thus the second inequality in (3.32a) is valid when $n=m$ too. Therefore, the proof of (3.32) is completed by the method of mathematical induction under the choice of $h_0=\min\{h_1,h_2,h_3\}$ and $\tau_0=\min\{1/8,\tau_2,\tau_3\}$. □

References

1. Ablowitz, M.J., Kruskal, M.D., Ladik, J.F.: Solitary wave collisions. SIAM J. Appl. Math. **36**, 428–437 (1979)
2. Adomian, G.: Nonlinear Klein-Gordon equation. Appl. Math. Lett. **9**, 9–10 (1996)
3. Baǐnov, D.D., Minchev, E.: Nonexistence of global solutions of the initial-boundary value problem for the nonlinear Klein-Gordon equation. J. Math. Phys. **36**, 756–762 (1995)
4. Bao, W., Cai, Y.: Optimal error estimates of finite difference methods for the Gross-Pitaevskii equation with angular momentum rotation. Math. Comput. (2011) (in press)
5. Bao, W., Dong, X., Xin, J.: Comparisons between sine-Gordon equation and perturbed nonlinear Schrödinger equations for modeling light bullets beyond critical collapse. Physica D **239**, 1120–1134 (2010)
6. Bao, W., Sun, F.F.: Efficient and stable numerical methods for the generalized and vector Zakharov system. SIAM J. Sci. Comput. **26**, 1057–1088 (2005)
7. Bao, W., Li, X.G.: An efficient and stable numerical method for the Maxwell-Dirac system. J. Comput. Phys. **199**, 663–687 (2004)
8. Bao, W., Yang, L.: Efficient and accurate numerical methods for the Klein-Gordon-Schrödinger equations. J. Comput. Phys. **225**, 1863–1893 (2007)
9. Brenner, P., van Wahl, W.: Global classical solutions of nonlinear wave equations. Math. Z. **176**, 87–121 (1981)
10. Browder, F.E.: On nonlinear wave equations. Math. Z. **80**, 249–264 (1962)
11. Cao, W., Guo, B.: Fourier collocation method for solving nonlinear Klein-Gordon equation. J. Comput. Phys. **108**, 296–305 (1993)
12. Cohen, D., Hairer, E., Lubich, Ch.: Conservation of energy, momentum and actions in numerical discretizations of non-linear wave equations. Numer. Math. **110**, 113–143 (2008)
13. Davydov, A.S.: Quantum Mechanics, 2nd edn. Pergamon, Oxford (1976)
14. Deeba, E.Y., Khuri, S.A.: A decomposition method for solving the nonlinear Klein-Gordon equation. J. Comput. Phys. **124**, 442–448 (1996)
15. Duncan, D.B.: Symplectic finite difference approximations of the nonlinear Klein-Gordon equation. SIAM J. Numer. Anal. **34**, 1742–1760 (1997)
16. Gautschi, W.: Numerical integration of ordinary differential equations based on trigonometric polynomials. Numer. Math. **3**, 381–397 (1961)
17. Ginibre, J., Velo, G.: The global Cauchy problem for the nonlinear Klein-Gordon equation. Math. Z. **189**, 487–505 (1985)
18. Ginibre, J., Velo, G.: The global Cauchy problem for the nonlinear Klein-Gordon equation–II. Ann. Inst. H. Poincaré Anal. Non Linéaire **6**, 15–35 (1989)
19. Glassey, R.: On the asymptotic behavior of nonlinear wave equations. Trans. Am. Math. Soc. **182**, 187–200 (1973)
20. Glassey, R., Tsutsumi, M.: On uniqueness of weak solutions to semi-linear wave equations. Commun. Partial Differ. Eqn. **7**, 153–195 (1982)
21. Grimm, V.: A note on the Gautschi-type method for oscillatory second-order differential equations. Numer. Math. **102**, 61–66 (2005)
22. Grimm, V.: On error bounds for the Gautschi-type exponential integrator applied to oscillatory second-order differential equations. Numer. Math. **100**, 71–89 (2005)
23. Grimm, V.: On the use of the Gautschi-type exponential integrator for wave equation. In: Berm'udez de Castro, A., G'omez, D., Quintela, P., Salgado, P. (eds.) Numerical Mathematics and Advanced Applications (ENUMATH2005), pp. 557–563. Springer, Berlin (2006)
24. Hairer, E., Lubich, Ch., Wanner, G.: Geometric Numerical Integration. Springer, Berlin (2002)
25. Hochbruck, M., Lubich, Ch.: A Gautschi-type method for oscillatory second-order differential equations. Numer. Math. **83**, 402–426 (1999)
26. Hochbruck, M., Ostermann, A.: Exponential integrators. Acta Numer. **19**, 209–286 (2000)
27. Ibrahim, S., Majdoub, M., Masmoudi, N.: Global solutions for a semilinear, two-dimensional Klein-Gordon equation with exponential-type nonlinearity. Commun. Pure Appl. Math. **59**, 1639–1658 (2006)
28. Jiménez, S., Vázquez, L.: Analysis of four numerical schemes for a nonlinear Klein-Gordon equation. Appl. Math. Comput. **35**, 61–94 (1990)

29. Kosecki, R.: The unit condition and global existence for a class of nonlinear Klein-Gordon equations. J. Differ. Equ. **100**, 257–268 (1992)
30. Li, S., Vu-Quoc, L.: Finite difference calculus invariant structure of a class of algorithms for the nonlinear Klein-Gordon equation. SIAM J. Numer. Anal. **32**, 1839–1875 (1995)
31. Machihara, S.: The nonrelativistic limit of the nonlinear Klein-Gordon equation. Funkcial. Ekvac. **44**, 243–252 (2001)
32. Machihara, S., Nakanishi, K., Ozawa, T.: Nonrelativistic limit in the energy space for nonlinear Klein-Gordon equations. Math. Ann. **322**, 603–621 (2002)
33. Masmoudi, N., Nakanishi, K.: From nonlinear Klein-Gordon equation to a system of coupled nonliner Schrödinger equations. Math. Ann. **324**, 359–389 (2002)
34. Morawetz, C., Strauss, W.: Decay and scattering of solutions of a nonlinear relativistic wave equation. Commun. Pure Appl. Math. **25**, 1–31 (1972)
35. Najman, B.: The nonrelativistic limit of the nonlinear Klein-Gordon equation. Nonlinear Anal. **15**, 217–228 (1990)
36. Najman, B.: The Nonrelativistic Limit of Klein-Gordon and Dirac Equations. Differential Equations with Applications in Biology, Physics, and Engineering (Leibnitz, 1989). Lecture Notes in Pure and Applied Mathematics, vol. 133. pp. 291–299. Dekker, New York (1991)
37. Nakamura, M., Ozawa, T.: The Cauchy problem for nonlinear Klein-Gordon equations in the Sobolev spaces. Publ. Res. Inst. Math. Sci. **37**, 255–293 (2001)
38. Pachpatte, B.G.: Marcel Inequalities for Finite Difference Equations. Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker Inc., New York (2002)
39. Pascual, P.J., Jiménez, S., Vázquez, L.: Numerical Simulations of a Nonlinear Klein-Gordon Model. Applications. Computational Physics (Granada, 1994). Lecture Notes in Physics, vol. 448. pp. 211–270. Springer, Berlin (1995)
40. Pecher, H.: Nonlinear small data scattering for the wave and Klein-Gordon equation. Math. Z. **185**, 261–270 (1984)
41. Sakurai, J.J.: Advanced Quantum Mechanics. Addison Wesley, New York (1967)
42. Segal, I.E.: The global Cauchy problem for a relativistic scalar field with power interaction. Bull. Soc. Math. Fr. **91**, 129–135 (1963)
43. Shatah, J.: Normal forms and quadratic nonlinear Klein-Gordon equations. Commun. Pure Appl. Math. **38**, 685–696 (1985)
44. Simon, J.C.H., Taflin, E.: The Cauchy problem for non-linear Klein-Gordon equations. Commun. Math. Phys. **152**, 433–478 (1993)
45. Smith, G.D.: Numerical Solution of Partial Differential Equations. Oxford University Press, London (1965)
46. Shen, J., Tang, T.: Spectral and High-Order Methods with Applications. Science Press, Beijing (2006)
47. Strauss, W.: Decay and asymptotics for $\Box u = f(u)$. J. Funct. Anal. **2**, 409–457 (1968)
48. Strauss, W., Vázquez, L.: Numerical solution of a nonlinear Klein-Gordon equation. J. Comput. Phys. **28**, 271–278 (1978)
49. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems. Springer, Berlin (1997)
50. Tourigny, Y.: Product approximation for nonlinear Klein-Gordon equations. IMA J. Numer. Anal. **9**, 449–462 (1990)
51. Tsutsumi, M.: Nonrelativistic approximation of nonlinear Klein-Gordon equations in two space dimensions. Nonlinear Anal. **8**, 637–643 (1984)